BEST AVAILABLE COPY

REMARKS

Claims 31-35 were previously pending and are amended herein. Claims 36-38 have been added. Support for the newly added claims can be found, for example, in paragraph 0137 of the application and in Figure 17. No new matter has been added.

Rejections Under 35 U.S.C. §112, first paragraph (written description)

The Examiner rejected claims 33-35 as allegedly failing to meet the written description requirement. The Examiner stated that the claims encompass variants of gag, pol and env. In particular the Examiner notes that the claims encompass variants of HIV pol having reduced reverse transcriptase activity, reduced strand transfer activity or reduced RNaseH activity. In making this rejection, the Examiner cited: *Fiers v. Revel, Amgen, Inc. v. Chugai Pharmaceuticals Co. Ltd, Fiddes v. Baird* and *University of California v. Eli Lilly and Co.* Applicants respectfully traverse this rejection.

The Examiner acknowledges that the present application discloses an HIV pol sequence having a mutation that reduce reverse transcriptase activity, a mutation that reduces strand transfer activity and a mutations that reduces RNaseH activity. However, the Examiner argues that the disclosure of these mutations is inadequate to meet the written description requirement for the claimed genus of variants. Applicants disagree.

The USPTO Written Description Requirement Guidelines explain that the written description requirement can be met by a "show[ing] that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics … i.e., complete or partial structure, other physical characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics." *Guidelines for Examination of Patent Applications Under 35 U.S.C. 112 Paragraph 1, "Written Description" Requirement.* 66 Fed. Reg. 1099, 1106. In *Enzo Biochem, Inc. v. Gen-Probe Inc.*, the Court of Appeals for the Federal Circuit quoted this portion of the Guidelines with approval stating that "we are persuaded by the Guidelines on this point and adopt the PTO's applicable standard for determining compliance with the written description requirement." 296 F.3d 1261 (Fed. Cir.

2002). Thus, whether a disclosure meets the written description required for a claimed invention, e.g., a claimed protein or variant protein depends, in part on the knowledge of those skilled in the art regarding the claimed protein or protein variant. For example, the existence of a known correlation between structure and function is a relevant factor in assessing whether the written description requirement has been met.

The HIV genome and the proteins expressed by HIV are extremely well studied. In particular, the genomes of many different HIV strains from a variety of clades have been sequenced in whole or in part. This is true today and it was true as of the priority date of the present application. In addition, the structure and function of many HIV proteins have been studied in detail. HIV Pol has been particularly intensively studied because it includes the protease and reverse transcriptase functions that are the targets of the major therapeutic treatments for AIDS.

The present application explains that GenBank® and various other sequence databases, such as that found on the internet at http://hiv-web.lanl.gov provide HIV sequences (see paragraph 0074 of the present specification). As of the priority date of the present application hundreds of different HIV strains had been sequenced, in whole or in part. The "HIV Sequences Compendium" is published annually by the Theoretical Biology and Biophysics Group, Los Alamos National Laboratory. This compendium provides the sequence of numerous HIV genomes and an even larger number can be found at a related web site (http://hiv-web.lanl.gov). Applicants have enclosed pages i-viii and pages 457 to 527 of the "HIV Sequences Compendium 2000", Kuiken et al., eds., published by the Theoretical Biology and Biophysics Group, Los Alamos National Laboratory (Exhibit A). This version of the Compendium includes HIV sequences published up until the end of 2000. Pages 457 to 527 of the "HIV Sequences Compendium 2000" provide an alignment of the amino acid sequence of the HIV-1 proteins (along with a few SIV cpz proteins). As explained on page 457 of the Compendium, the envelope master sequence alignment alone includes 307 full-length sequences. The Compendium itself includes a representative selection of the alignments. The alignments for HIV pol begin on page 480. As the Examiner can see, Nearly 100 HIV pol sequences from

various clades are aligned with the HXB2 sequence (This is the HIV pol sequence used as the basis for the recombinant MVA described in Example 2 (paragraphs 0136-0154) of the present application). The conserved and variant amino acids are shown and consensus sequences for various clades are presented. Along the top of the pol alignment various functional domains and landmarks within pol, for example, the domains associated with RNaseH activity and integrase activity, are identified. The Compendium includes similar alignments for all of the other HIV proteins. The "HIV Sequences Compendium 2000" provides an alignment of a number of HIV pol amino acid sequences with the HIV HXB2 pol sequence. This is exactly the type of publicly available information that one can use to identify other HIV pol variants having a mutation that inhibits reverse transcriptase activity, RNaseH activity or strand transfer activity. It is apparent that many, many HIV protein sequences were known in the art as of the priority date of the present application and that alignment of the protein sequences permitted those in the art to identify conserved residues and functional domains.

Numerous mutations within HIV pol that inhibit reverse transcriptase activity, RNaseH activity or strand transfer activity are known to those skilled in the art. For example, Snyder et al. (*J. Virol.* 74:9668, 2000; Exhibit B) describes at mutation at amino acid 478 of HIV pol that interferes with strand transfer activity. Fan et al. (*Biochemistry* 35:9737, 1997; Exhibit C; under separate cover) Gao et al. (*J. Mol. Biol.* 277:559, 1998; Exhibit D; under separate cover) and Powell et al. (*J. Biol. Chem.* 2772:13262, 1997; Exhibit E; under separate cover) describe HIV mutations that reduce RNaseH activity. Finally, Boyer et al. (*Proc. Nat'l Acad. Sci.* 97:3056, 2000; Exhibit F) describes mutations that reduce reverse transcriptase activity. These references are just a sampling of the many publications describing mutations that inhibit the activity of HIV pol.

As noted above, the present specification identifies mutations in HIV pol that lead to inhibition of reverse transcriptase activity, inhibition of strand transfer activity or inhibition of RNaseH activity. These mutations are located at amino acids 185, 266 and 478 of HIV strain HXB2 pol respectively (see paragraphs 0136 and 0137 of the present application). This disclosure in the context of the very large number of publicly available HIV pol sequences and

HIV pol mutations described in the scientific literature, is adequate to meet the written description requirement

Given the large number of HIV sequences known to researchers, Applicants description of mutations in HIV HXB2 pol that lead to reduced reverse transcriptase activity, reduced strand transfer activity or reduced RNaseH is sufficient to meet the written description requirement of 35 U.S.C. §112, first paragraph for the present claims. In view of this, Applicants respectfully request that the Examiner withdraw this written description rejection under 35 U.S.C. §112, first paragraph.

Rejections Under 35 U.S.C. §112, first paragraph (new matter)

The Examiner rejected claims 33-35 as allegedly containing new matter. The Examiner argued that the phrases: "reducing reverse transcriptase activity", "reducing strand transfer activity", and "reducing RNaseH activity" in claims 33-35, respectively, constitute new matter. Specifically, the Examiner stated that these phrases do not appear in the specification. The exact phrases used in a claim do not need to appear in the specification. *Cordis Corp. v. Medtronic Ave, Inc.*, 339 F.3d 1352 (Fed. Cir. 2003) ("The disclosure as originally filed, does not, however, have to provide *in haec verba* support for the claimed subject matter at issue"). However, Applicants have amended claims 33-35 to recite: "inhibits reverse transcriptase activity", "inhibit strand transfer activity", and "inhibits RNaseH activity", respectively. Support for these limitations can be found, for example, in paragraph 0137 of the present specification.

In view of the forgoing, Applicants respectfully request that the Examiner withdraw this new matter rejection under 35 U.S.C. §112, first paragraph.

Rejections Under 35 U.S.C. §112, second paragraph

The Examiner rejected claims 33-35 as allegedly indefinite. The Examiner argued that the terms "reverse transcriptase activity", stand transfer activity" and RNaseH activity" are indefinite. The Examiner ask which activity the various phrases refer to and in what biological context.

"If one skilled in the art would understand the bounds of the claim when read in the light of the specification, then the claim satisfies section 112, paragraph 2." *Miles Labs., Inc. v. Shandon, Inc.*, 997 F.2d 870, 875 (Fed. Cir. 1993). It is Applicants' position that the phrases "reverse transcriptase activity", stand transfer activity" and "RNaseH activity" in the context of the present application would be understood by those of ordinary skill in the art. Moreover, as discussed at length above, HIV pol had been well studied at the time of priority date of the present application. Each of the activities mentioned in the claims had been characterized and mutations interfering with these activities had been identified. Thus, those skilled in art well-understood the various activities possessed by HIV pol, and the phrases "reverse transcriptase activity", stand transfer activity" and "RNaseH activity" in the context of the present application would be clear to those of ordinary skill in the art.

In view of the forgoing, Applicants respectfully request that these rejections under 35 U.S.C. §112, second paragraph be withdrawn.

Rejections Under 35 U.S.C. §103

The Examiner rejected claims 31-35 as obvious in view of Kent et al. (*J. Virol.* 72:10181, 1998; "Kent") taken with Small et al. According to the Examiner, Kent discloses "priming with DNA and boosting with a recombinant Fowlpox virus wherein said DNA and recombinant Fowlpox encode (express) the *gag*, *pol* and *env* genes of HIV-1" and Small discloses "that MVA constitutes a safe version of recombinant pox virus". The Examiner argued that it would have been obvious to "use the recombinant MVA disclosed by Small et al. in the method described by Kent et al, in order to take advantage of the increased safety associated with the use of MVA virus in immunization protocols."

Applicants respectfully traverse this rejection. The MVA employed in the presently claimed methods expresses: HIV Gag, HIV Pol lacking the integrase domain, HIV gp120 and HIV gp41, lacking or all or part of its cytoplamic domain."

Kent describes two different recombinant fowlpox virus. One recombinant virus expresses HIV gag and HIV pol. The other recombinant poxvirus expresses HIV env. These are

described on page 10181 in the second paragraph under the heading "Recombinant Poxvirus". Thus, none of the recombinant poxvirus used by Kent expressed HIV gag, HIV pol and any portion of HIV env. In addition, the HIV pol-expressing poxvirus used by Kent apparently expressed intact HIV pol, not HIV pol lacking the integrase domain, as required by the present claims. Even if Small et al. suggests that MVA is a useful poxvirus for expressing heterologous genes, the combination of Kent and Small et al. would not lead one to the recombinant MVA used in the present methods. Thus, the cited references, no matter how combined, do not teach or suggest the limitations of claim 31. Accordingly, the cited references do not render claim 31 obvious.

Regarding claims 33-35, which specify that the HIV pol expressed by the recombinant MVA lacks certain activities, the Examiner states that "it is deemed (in the absence of evidence to the contrary) that the recited genes would not function properly in the context of the claimed recombinant virus due to the lack of normal accessory proteins, enhancer etc." Applicants respectfully request that the Examiner provide a more complete explanation of the rejection of these claims. The HIV pol expressed is intended to be an antigen, not a functional HIV pol. Indeed, mutations can be desirable to reduce enzymatic activity that supports HIV replication. In any event, claims 33-35 all depend from claim 31. As explained above, the cited references, no matter how combined do not teach or suggest the limitations of claim 31. According, the cited references cannot render any of the dependent claims (claims 32-38) obvious.

In view of the forgoing, Applicants respectfully request that the rejections under 35 U.S.C. §103 be withdrawn.

Enclosed is a Petition for Extension of Time with the appropriate fee. Please apply any other charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: 20 MARCH 2006

Anita L. Meiklejohn, Ph.D.
Reg. No. 35,283

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

21290482.doc

OIPE
IAP17
MAR 23 2006
PATENT & TRADEMARK OFFICE

CONTENTS

More detailed tables of contents can be found within the various parts of the compendium.

Acknowledgments ii
Introduction iii
Maps of HIV and SIV Genomes iv
Landmarks of the Genome v

PART I. REVIEWS

Tat, a Novel Regulator of HIV Transcription and Latency 2
Jonathan Karn
An Overview of HIV-1 Co-receptor Function and Its Inhibitors 19
Emmanuel G. Cormier and Tatjana Dragic
An Overview of the Molecular Phylogeny of Lentiviruses 35
Brian T. Foley
Sequence Homology Search Tools on the World Wide Web 44
Ian Holmes
Recombinant HIV Sequences: Their Role in the Global Epidemic 54
Martine Peeters
Reagents for HIV/SIV Vaccine Studies 73
Rama Thakallapally and Carla Kuiken
Clinical Trials of HIV Vaccines 82
Barney S. Graham
Mutations in Retroviral Genes Associated with Drug Resistance 106
Urvi Parikh, Jennifer Hammond, Charles Calef, Brendan Larder, Raymond Schinazi, John W. Mellors

PART II. HIV-1/SIVcpz NUCLEOTIDE ALIGNMENTS

Contents 163
Introduction 164
Table of HIV-1/SIVcpz Sequences in the Nucleotide Alignment 167
Notes on full-length HIV-1/SIVcpz Sequences in the Nucleotide Alignment 170
Nucleotide Alignment of HIV-1/SIVcpz Complete Genomes 188

PART III. HIV-1/HIV-2/SIV NUCLEOTIDE ALIGNMENTS

Contents 355
Introduction 355
Table of Primate Lentiviruses Sequences in the Nucleotide Alignment 357
Nucleotide Alignment of Primate Lentiviral Complete Genomes 358

PART IV. HIV-1/SIVcpz AMINO ACID ALIGNMENTS

Contents 457
Introduction
Table of HIV-1/SIVcpz sequences in the Amino Acid Alignments 460
Amino Acid Alignments of HIV-1/SIVcpz 470

PART V. HIV-2/SIV AMINO ACID ALIGNMENTS

Contents 531
Introduction 532
Table of HIV-2/SIV sequences in the Amino Acid Alignments 533
Amino Acid Alignments of HIV-2/SIV 537

PART VI. Other SIV AMINO ACID ALIGNMENTS

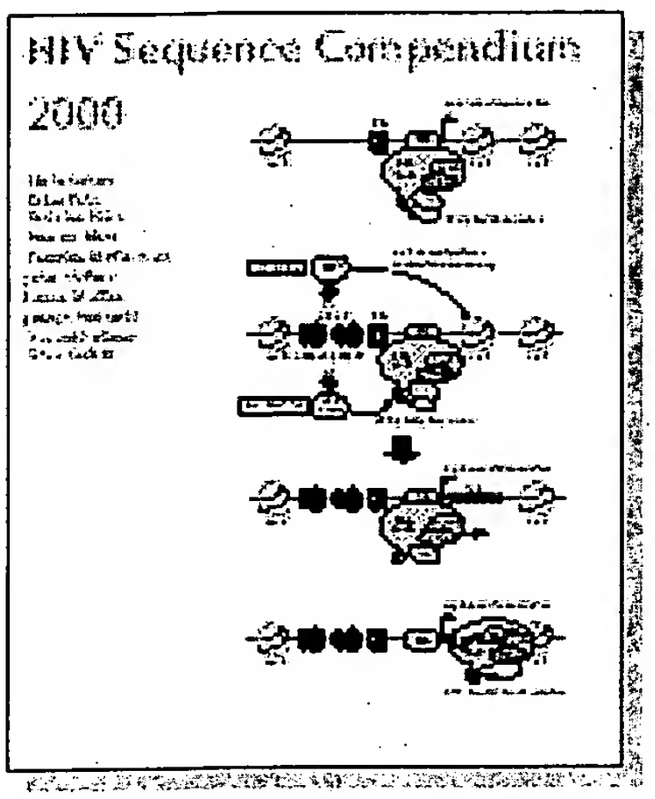
Contents 567
Table of other SIV sequences in the Amino Acid Alignments 568
Amino Acid Alignments of Other SIV 570

ACKNOWLEDGMENTS

The HIV Sequence Database and Analysis Project is funded by the Vaccine and Prevention Research Program of the AIDS Division of the National Institute of Allergy and Infectious Diseases (Dr. James Bradac, Project Officer) through an interagencyagreement with the U.S. Department of Energy.

We thank our editors, the many researchers who have made their sequences available prior to publication, and authors who help by contributing to our review section.

The Cover



A schematic representation of the activation mechanism of latent proviruses by NF-κB and Tat during T-cell activation. From: Karn J, Tat, a novel regulator of HIV transcription and latency, Page 2 of this Compendium.

Citing this publication

We have simplified the name of this, our annual publication. Formerly known as "Human Retroviruses and AIDS" it should now be cited simply as *HIV Sequence Compendium 2000*, Kuiken C, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, Mullins J, Wolinsky S, and Korber B, editors. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.

Introduction

This Compendium is an annual printed summary of the data contained in the HIV sequence database. In these compendia we try to present a judicious selection of the data in such a way that it is of maximum utility to HIV researchers. Traditionally, we present the sequence data themselves in the form of alignments: a comprehensive alignment of a selection of all full-length genomes the database contains (a lot of LAI-like sequences, for example, have been omitted because they are so similar that they bias the alignment) of HIV-1/SIVcpz (Section I) and a combined HIV-1/HIV-2/SIV whole genome alignment (Section II); amino acid alignments for HIV-1/SIV-cpz, HIV-2/SIV, and SIVagm. The HIV-2/SIV and SIVagm amino acid alignments are separate because the genetic distances between these groups are so great that presenting them in one alignment would make them very elongated because of the large number of gaps that have to be inserted. As always, tables with extensive background information gathered from the literature accompany the whole genome alignments.

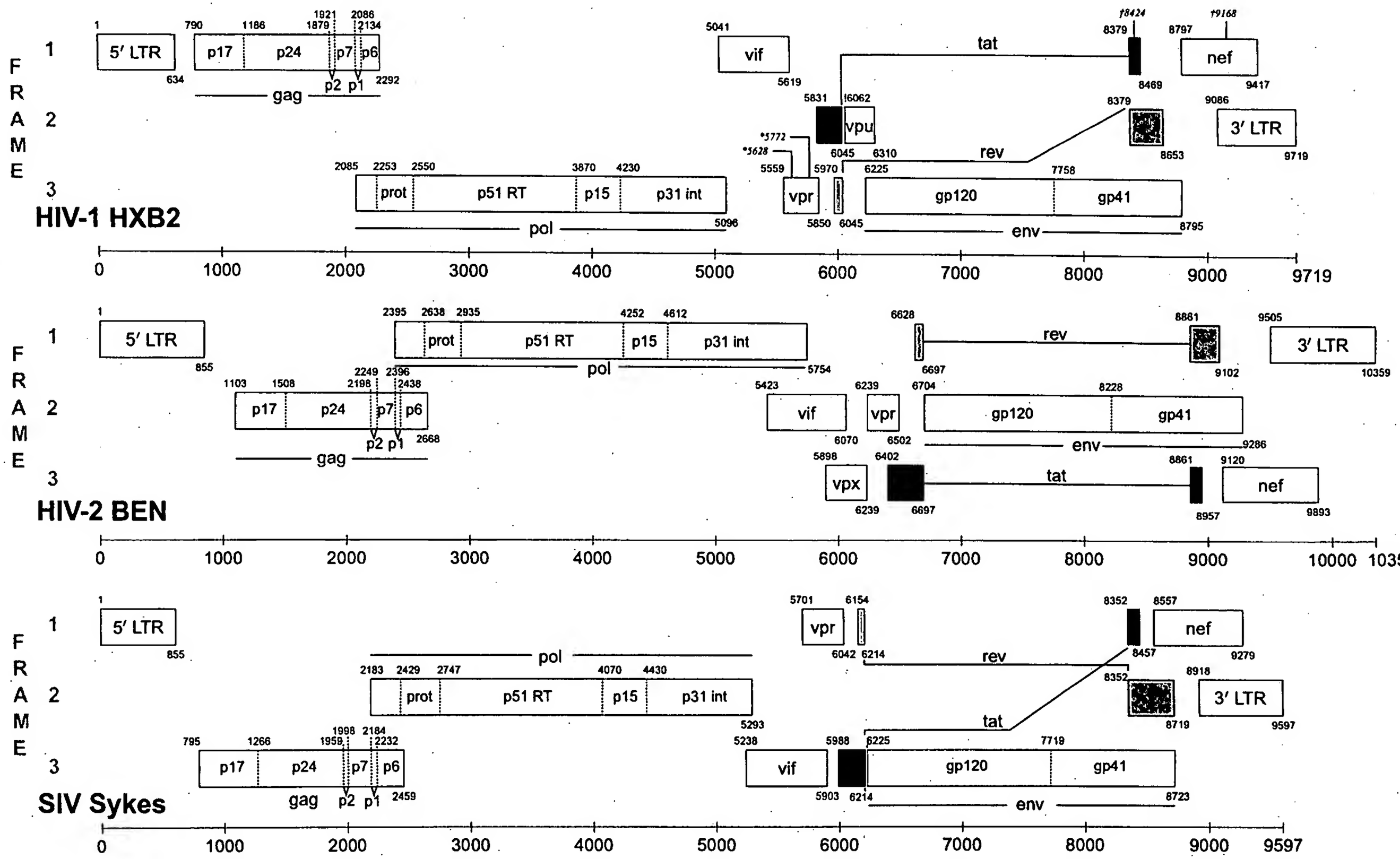
The collection of whole-gene sequences in the database is now large enough that we have abundant representation of most subtypes (excluding H and J). For most other subtypes, and especially for subtype B, a large number of sequences that span entire genes were not included in the printed alignments to conserve space. A more complete version of all alignments is available on our website, http://hiv-web.lanl.gov Importantly, all these alignments have been edited to include only one sequence per person, based on phylogenetic trees that were created for all of them, as well as the literature. At the request of many users, we have re-inserted the consensus sequences for each subtype, unless there are fewer than five sequences representing it. In the alignments we have also included the 'Circulating Recombinant Forms', mosaic genomes that have epidemiological significance (see the nomenclature chapter for more on CRFs). Finally, for all amino acid alignments we have decided to combine the annotation tables into one, because of the increasing redundancy in the separate tables. In addition to sequence information (accession numbers, references) the new table lists which regions of the sequence are represented in the alignments.

We have made an effort to bring the HIV-2/SIV and SIVagm alignments up-to-date as well. We have created an entirely new HIV-1/HIV-2/SIV alignment that is much improved over the previous version; it can be accessed via our website. Because of the frequency of redundant information, we have decided to merge the gene tables into one large table for each alignment section; we hope you will find these tables easy to use.

In the Reviews section, along with the previously mentioned chapter on HIV nomenclature, you will find a very clearly written and concise overview of the functions of TAT; a review of lentivirus phylogeny and evolution; an overview of recombinants and their role in the epidemic; a concise and lucid chapter on protein search tools on the internet; an overview of HIV-1 coreceptors and coreceptor inhibitors; and a very thorough review on progress in SIV and HIV vaccines. In addition, we present updated versions of the customary reviews of coreceptor usage, drug resistance, and SIV/SHIV vaccine reagents. Reprints of all reviews are available from our website in the form of both HTML and PDF files.

As always, we are open to complaints and suggestions for improvement. With the effort that goes into producing these volumes, we sincerely hope they will be widely used by the research community. Inquiries and comments regarding the Compendium should be addressed to:

Dr. Carla Kuiken
Theoretical Division, T-10, MS K710, LANL, Los Alamos, NM 87545
Ph: (505)-665—6463; fax: (505)-665—3493; e-mail: kuiken@t10.lanl.gov



Landmarks of the HIV-1, HIV-2, and SIV genomes. The gene start, indicated by the small number in the upper left corner of each rectangle normally records the position of the `a` in the `atg` start codon for that gene while the number in the lower right records the last position of the stop codon. For *pol*, the start is taken to be the first `t` in the sequence `ttttttag` which forms part of the stem loop that potentiates ribosomal slippage on the RNA and a resulting –1 frameshift and the translation of the gag-pol polyprotein. The *tat* and *rev* spliced exons are shown as shaded rectangles. In HXB2, **5628* and **5772* mark positions of frameshifts in the `vpr` gene; !6062 indicates a defective `acg` start codon in *vpu*; *†8424*, and *†9168* mark premature stop codons in *tat* and *nef*. See Korber et al., Numbering Positions in HIV Relative to HXB2CG, in *Human Retroviruses and AIDS, 1998* p. 102. Available from `http://hiv-web.lanl.gov/HTML/reviews/HXB2.html`.

HIV/SIV PROTEINS

NAME	SIZE	FUNCTION	LOCALIZATION
Gag MA	p17	membrane anchoring; env interaction; nuclear transport of viral core. (myristylated protein)	virion
CA	p24	core capsid	virion
NC	p7	nucleocapsid, binds RNA	virion
	p6	binds Vpr	virion
Protease (PR)	p15	gag/pol cleavage and maturation	virion
Reverse transcriptase (RT), RNase H	p66 p51	reverse transcription, RNase H activity	virion
Integrase (IN)		DNA provirus integration	virion
Env	gp120 gp41	external viral glycoproteins bind to CD4 and secondary receptors	plasma membrane, virion envelope
Tat	p16/p14	viral trancriptional transactivator	primarily in nucleolus/nucleus
Rev	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nucleolus/nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity	cytoplasm (cytosol, membranes) virion
Vpr	p10-15	promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M	virion, nucleus (nuclear membrane?)
Vpu	p16	promotes extracellular release of viral particles; degrades CD4 in the ER; (phosphoprotein only in HIV-1 and SIVcpz)	integral membrane protein
Nef	p27-p25	CD4 and class I downregulation (myristylated protein)	plasma membrane, cytoplasm (virion?)
Vpx	p12-16	vpr homolog (not in HIV-1, only in HIV-2 and SIV)	virion (nucleus?)

LANDMARKS:

HIV GENOMIC STRUCTURAL ELEMENTS

LTR Long terminal repeat, the DNA sequence flanking the genome of integrated proviruses. It contains important regulatory regions, especially those for transcription initiation and polyadenylation.

TAR Target sequence for viral transactivation, the binding site for Tat protein and for cellular proteins; consists of approximately the first 45 nucleotides of the viral mRNAs in HIV-1 (or the first 100 nucleotides in HIV-2 and SIV.) TAR RNA forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function.

RRE Rev responsive element, an RNA element encoded within the env region of HIV-1. It consists of approximately 200 nucleotides (positions 7327 to 7530 from the start of transcription in HIV-1, spanning the border of gp120 and gp41). The RRE is necessary for Rev function; it contains a high affinity site for Rev; in all, approximately seven binding sites for Rev exist within the RRE RNA. Other lentiviruses (HIV-2, SIV, visna, CAEV) have similar RRE elements in similar locations within env, while HTLVs have an analogous RNA element (RXRE) serving the same purpose within their LTR; RRE is the binding site for Rev protein, while RXRE is the binding site for Rex protein. RRE (and RXRE) form complex secondary structures, necessary for specific protein binding.

CRS Cis-acting repressive sequences postulated to inhibit structural protein expression in the absence of Rev. One such site was mapped within the pol region of HIV-1. The exact function has not been defined; splice sites have been postulated to act as CRS sequences.

INS Inhibitory/Instability RNA sequences found within the structural genes of HIV-1 and of other complex retroviruses. Multiple INS elements exist within the genome and can act independently; one of the best characterized elements spans nucleotides 414 to 631 in the gag region of HIV-1. The INS elements have been defined by functional assays as elements that inhibit expression posttranscriptionally. Mutation of the RNA elements was shown to lead to INS inactivation and up regulation of gene expression.

GENES AND GENE PRODUCTS

GAG The genomic region encoding the capsid proteins (group specific antigens). The precursor is the p55 myristylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCapsid), and p6 proteins, by the viral protease. Gag associates with the plasma membrane where the virus assembly takes place. The 55 kDa Gag precursor is called assemblin to indicate its role in viral assembly.

POL The genomic region encoding the viral enzymes protease, reverse transcriptase and integrase. These enzymes are produced as a Gag-pol precursor polyprotein, which is processed by the viral protease; the Gag-pol precursor is produced by ribosome frameshifting at the C-terminus of gag.

ENV Viral glycoproteins produced as a precursor (gp160) which is processed to give a noncovalent complex of the external glycoprotein gp120 and the transmembrane glycoprotein gp41. The mature gp120-gp41 proteins are bound by non-covalent interactions and are associated as a trimer on the cell surface. A substantial amount of gp120 can be found released in the medium. gp120 contains the binding site for the CD4 receptor, and the seven transmembrane domain chemokine receptors that serve as co-receptors for HIV-1.

TAT Transactivator of HIV gene expression. One of two essential viral regulatory factors (Tat and Rev) for HIV gene expression. Two forms are known, Tat-1 exon (minor form) of 72 amino acids and Tat-2exon (major form) of 86 amino acids. Low levels of both proteins are found in persistently infected cells. Tat has been localized primarily in the nucleolus/nucleus by immunofluorescence. It acts by binding to the TAR RNA element and activating transcription

initiation and/or elongation from the LTR promoter. It is the first eukaryotic transcription factor known to interact with RNA rather than DNA and may have similarities with prokaryotic anti-termination factors. Extracellular Tat can be found and can be taken up by cells in culture.

REV The second necessary regulatory factor for HIV expression. A 19 kD phosphoprotein, localized primarily in the nucleolus/nucleus, Rev acts by binding to RRE and promoting the nuclear export, stabilization and utilization of the viral mRNAs containing RRE. Rev is considered the most functionally conserved regulatory protein of lentiviruses. Rev cycles rapidly between the nucleus and the cytoplasm.

VIF Viral infectivity factor, a basic protein of typically 23 kD. Promotes the infectivity but not the production of viral particles. In the absence of Vif the produced viral particles are defective, while the cell-to-cell transmission of virus is not affected significantly. Found in almost all lentiviruses, Vif is a cytoplasmic protein, existing in both a soluble cytosolic form and a membrane-associated form. The latter form of Vif is a peripheral membrane protein that is tightly associated with the cytoplasmic side of cellular membranes. Some recent observations suggest that Vif functions late in replication to modulate assembly, budding, and/or maturation the N-terminal half of Vif (N'-Vif) specifically interacts with viral protease.

VPR Vpr (viral protein R) is a 96-amino acid (14 kd) protein, which is incorporated into the virion. It interacts with the p6 gag part of the Pr55 gag precursor. Vpr detected in the cell is localized to the nucleus. Proposed functions for Vpr include the targeting the nuclear import of preintegration complexes, cell growth arrest, transactivation of cellular genes, and induction of cellular differentiation. It is found in HIV-1, HIV-2, SIVmac and SIVmnd. It is homologous to the vpx protein.

VPU Vpu (viral protein U) is unique to HIV-1 and SIVcpz, a close relative of HIV-1. There is no similar gene in HIV-2 or other SIVs. Vpu is a 16-kd (81-amino acid) type I integral membrane protein with at least two different biological functions: (a) degradation of CD4 in the endoplasmic reticulum, and (b) enhancement of virion release from the plasma membrane of HIV-1-infected cells. Env and Vpu are expressed from a bicistronic mRNA. Vpu probably possesses an N-terminal hydrophobic membrane anchor and a hydrophilic moiety. It is phosphorylated by casein kinase II at positions Ser52 and Ser56. Vpu is involved in env maturation and is not found in the virion. Vpu has been found to increase susceptibility of HIV-1 infected cells to Fas killing.

NEF A multifunctional 27-kd myristylated protein produced by an ORF located at the 3′ end of the primate lentiviruses. Other forms of Nef are known, including nonmyristylated variants. Nef is predominantly cytoplasmic and associated with the plasma membrane via the myristyl residue linked to the conserved second amino acid (Gly). Nef has also been identified in the nucleus and found associated with the cytoskeleton in some experiments. One of the first HIV proteins to be produced in infected cells, it is the most immunogenic of the accessory proteins. The nef genes of HIV and SIV are dispensable *in vitro*, but are essential for efficient viral spread and disease progression *in vivo*. Nef is necessary for the maintenance of high virus loads and for the development of AIDS in macaques, and viruses with defective Nef have been detected in some HIV-1 infected long term survivors. Nef downregulates CD4, the primary viral receptor, and MHC class I molecules, and these functions map to different parts of the protein. Nef interacts with components of host cell signal transduction and clathrin-dependent protein sorting pathways. It increases viral infectivity. Nef contains PxxP motifs that bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of HIV but not for the downregulation of CD4.

VPX A virion protein of 12 kD found only in HIV-2/SIVmac/SIVsm and not in HIV-1 or SIVagm. This accessory gene is a homolog of HIV-1 vpr, and HIV-2/SIV carry both vpr and vpx. Vpx function in relation to xpr is not fully elucidated; both are incorporated into virions at levels comparable to gag proteins through interactions with Gag p6. Vpx is necessary for efficient replication of SIV in PBMCs. Progression to AIDS and death in SIV-infected animals can occur in the absence of Vpr or Vpx. Double mutant virus lacking both vpr and vpx was attenuated,

whereas the single mutants were not, suggesting a redundancy in the function of Vpr and Vpx related to virus pathogenicity.

STRUCTURAL PROTEINS/VIRAL ENZYMES The products of gag, pol and env genes, which are essential components of the retroviral particle.

REGULATORY PROTEINS Tat and Rev proteins of HIV/SIV and Tax and Rex proteins of HTLVs. They modulate transcriptional and posttranscriptional steps of virus gene expression and are essential for virus propagation.

ACCESSORY OR AUXILIARY PROTEINS Additional virion and non-virion- associated proteins produced by HIV/SIV retroviruses: Vif, Vpr, Vpu, Vpx, Nef. Although the accessory proteins are in general not necessary for viral propagation in tissue culture, they have been conserved in the different isolates; this conservation and experimental observations suggest that their role *in vivo* is very important. Their functional importance continues to be elucidated.

COMPLEX RETROVIRUSES Retroviruses regulating their expression via viral factors and expressing additional proteins (regulatory and accessory) essential for their life cycle.

IV

HIV-1/SIVcpz proteins

Introduction 457
Table of HIV-1/SIVcpz protein alignments 460
Gag 470
Pol 480
Vif 498
Vpr 502
Tat 504
Rev 506
Vpu 508
Env 510
Nef 526

Construction of HIV-1/SIVcpz protein alignments

The number of full-length gene sequences is still growing rapidly for all genes. The envelope master alignment now contains 307 full-length sequences. For the purposes of the printed alignments, we have had to limit the number of sequences dramatically. Here we list the criteria we have followed to make the selection.

First, we have decided to end the supremacy of the B clade sequences. More than half (198, to be precise) of the full-length envelope sequences are still subtype B, though the contribution of other subtypes is increasing. We have tried to balance the number of representatives of all subtypes in these alignments. For this, we had to make a heavy selection on subtype B sequences. We have tried to include as many "classical" sequences as possible. A lot of follow-up work has been done based on lab strains such as HXB2, MN, SF2, and JR-CSF/JR-FL, so these strains are included in the alignments. Furthermore, within subtype B we have tried to represent sequences from diverse geographical origins, so as to represent a broad spectrum of variants. In the case of subtype B, this means that we have included African, Asian and Brazilian variants along with the "Western" strains. For sequences from non-B subtypes, we have selected a few representative sequences from each dataset, again with an eye on maintaining geographical diversity. When possible we have left all representatives of group O in the alignment, as these sequences are much more genetically diverse that the subtypes.

Explanation of Symbols in Alignments

Symbol	Meaning
Alignment symbols	
? in consensus	no majority-rule consensus could be determined at this position
x	nucleotide missing from codon
#	frameshift, or codon contains N or illegal character
$	stop codon
Annotation symbols	
|- -|	domain boundaries
/	protein start point
\	protein end point
\/	splice site or exon join
->	start of overlapping coding region
<-	end of overlapping coding region
*	cysteine
^^^ [NxS, NxT]	glycosylation site
^*^ [NCS, NCT]	glycosylation site with cysteine
CD4	residue critical for CD4 binding
cds	coding sequence (indicates regions where two proteins overlap; the overlapping proteins use two different reading frames)
MHR	major homology region
nls	nuclear localization signal
phos site	phosphorylation site
PKC	protein kinase C binding
Zn-motif	Zinc finger binding motif

Sources of Annotation in the Alignments

Protein	Annotation	Reference
Gag	phos site Ser (111)	Yu, J Biol Chem 270:4792 (1995)
Gag	MHR, (284-302)	Otteken, J Virol 70:3407 (1996)
Gag	CyPa (205-241)	Braaten, J Virol 70:4220 (1996)
Gag	vpr packaging domain LKSLFG, (489-494)	Lu, J Virol 69:6873 (1995) Kondo, J Virol 70:159 (1996)
Nef	myristylation, (1-7)	Huang, J Virol 69:93 (1995)
Nef	MHC downmodulation, PK recruitment (26-29)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	heart of CD4 binding site (55-56)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	acidic cluster, (60-64)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	(PxxP)3, (67-76)	Huang, J Virol 69:93 (1995)
Nef	PKC, (75-80)	Huang, J Virol 69:93 (1995)
Nef	polypurine tract, (89-97)	Huang, J Virol 69:93 (1995)
Nef	PAK binding, (103-105)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	Beta turn, (128-131)	Huang, J Virol 69:93 (1995)
Nef	PxxP, (145-148)	Huang, J Virol 69:93 (1995)
Nef	COP1 recruitment (152-153)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	AP recruitment, (162-163)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Nef	V-ATPase and Raf-1 binding, (172-173)	Piguet, p. 448 Human Retroviruses and AIDS (1999)
Vpr	alpha helix, (16-34)	Cornelissen, ARHR 13:247 (1997)
Vpr	H(S/F)RIG motifs, (71-82)	Macreadie, PNAS USA 92:2770 (1995)
Vpu	all annotations	Cornelissen, ARHR 13:247 (1997)
Vpr	LR domain, (60-82)	Wang, Gene 178:7 (1996)

Table 1: Table of HIV-1/SIVcpz protein Alignments

Name	Accession	Region	Author	Reference
A.DE.AF200476	AF200476	VIF	Kuhn, J	Unpublished
A.FR.HIV232956	AJ232956	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
A.GB.MA246	Y13718	ENV	Douglas, NW	J Mol Biol 273(1):122-149 (1997)
A.KE.AF233689	AF233689	VIF	Kuhn, J	Unpublished
A.KE.K89	L22943	ENV	Louwagie, J	J Virol 69(1):263-271 (1995)
A.KE.Q23	AF004885	ENV GAG NEF POL REV TAT VIF VPR VPU	Poss, M	Unpublished
A.RW.PVPI	L07082	ENV	Bex, F	Unpublished (1992)
A.SE.SE6594	AF069672	GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	AIDS 13(14):1819-1826 (1999)
A.SE.SE7253	AF069670	GAG POL REV TAT VIF VPR VPU	Carr, JK	AIDS 13(14):1819-1826 (1999)
A.SE.SE7535	AF069671	GAG POL REV TAT VIF VPR	Carr, JK	AIDS 13(14):1819-1826 (1999)
A.SE.SE8538	AF069669	GAG NEF POL REV TAT VIF VPR	Carr, JK	AIDS 13(14):1819-1826 (1999)
A.SE.SE8891	AF069673	GAG NEF REV TAT VIF VPR	Carr, JK	AIDS 13(14):1819-1826 (1999)
A.SE.UGSE8131	AF107771	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	-
A.UA.ukr970063	AF082486	ENV REV VPU	Liitsola, K	AIDS 12(14):1907-1919 (1998)
A.UG.92UG037	U51190	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 70(3):1651-1657 (1996)
A.UG.U13-2	X91354	VIF	Wieland, U	J Gen Virol 78:393-400 (1997)
A.UG.U455	M62320	ENV GAG NEF POL REV TAT VPR VPU	Oram, JD	ARHR 6(9):1073-1078 (1990)
A.UG.UG273A	L22957	REV TAT VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
A.UG.UG275A	L22951	REV TAT VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
A2.CD.97CDKFE4	AF286240	POL VIF VPR	Gao, F	ARHR 17(8):675-688(2001)
A2.CD.97CDKS10	AF286241	ENV REV TAT VIF VPR VPU	Gao, F	ARHR 17(8):675-688(2001)
A2.CD.97CDKTB48	AF286238	ENV GAG NEF REV TAT VIF VPR VPU	Gao, F	ARHR 17(8):675-688(2001)
A2.CY.94CY017-41	AF286237	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	ARHR 17(8):675-688(2001)

Table 1: cont.

Name	Accession	Region	Author	Reference
B.AU.MBC18	AF042102	ENV GAG POL REV VIF VPR VPU	Oelrichs, RB	ARHR 14(9):811-814 (1998)
B.AU.MBC200	AF042100	NEF TAT	Oelrichs, RB	ARHR 14(9):811-814 (1998)
B.AU.SC49	AF128998	GAG	Oelrichs, RB	Unpublished
B.AU.VH	AF146728	GAG REV VPR VPU	Oelrichs, R	Unpublished
B.CN.RL42	U71182	ENV GAG NEF POL REV TAT VIF VPR VPU	Graf, M	ARHR 14(3):285-288 (1998)
B.DE.D31	U43096	ENV GAG NEF POL REV TAT VIF VPR VPU	Kreutz, R	ARHR 8(9):1619-1629 (1992)
B.DE.HAN	U43141	ENV GAG NEF REV	Sauermann, U	ARHR 6(6):813-823 (1990)
B.ES.89SP061	AJ006287	ENV GAG NEF REV TAT VIF VPR VPU	Olivares, I	ARHR 14(18):1649-165 (1998)
B.FR.HXB2	K03455	ENV GAG NEF POL REV TAT VIF VPR VPU	Wong-Staal, F	Nature 313(6000):277-284 (1985)
B.FR.NE100	M58272	NEF	Delassus, S	J Virol 65:225-231 (1991)
B.FR.SWB884	M58206	NEF	Delassus, S	J Virol 65:225-231 (1991)
B.FR.vi02011A1H	AF143115	VIF	Hassaine, G	Virology 276(1):169-180 (2000)
B.GA.OYI	M26727	ENV GAG NEF POL REV TAT VIF VPR VPU	Huet, T	AIDS 3(11):707-715 (1989)
B.GB.CAM1	D10112	ENV GAG NEF POL REV TAT VIF VPR VPU	McIntosh, AA	Unpublished (1991)
B.GB.GB8	AJ271445	GAG	Farrar, GH	J Med Virol 34(2):104-113 (1991)
B.GB.I4663	Z68564	VPR	Kuiken, CL	J Gen Virol 77(Pt 4):783-792 (1996)
B.GB.I4663	Z68613	VPU	Kuiken, CL	J Gen Virol 77(Pt 4):783-792 (1996)
B.GB.MANC	U23487	GAG	Zhu, T	Nature 374(6522):503-504 (1995)
B.GB.WB	U36882	ENV	Douglas, NW	AIDS 10(1):39-46 (1996)
B.IN.HIVP35A	Y15122	NEF	Ahmad, KM	ARHR 14(16):1491-1493 (1998)
B.IT.B-IT-R5	AF147737	NEF	Catucci, M	J Med Virol 60(3):294-299 (2000)
B.JP.D70887	D70887	VIF	Tominaga, K	ARHR 12(16):1543-1549 (1996)
B.JP.ETR	D12582	ENV	Shimizu, H	Virology 189:534-546 (1992)
B.JP.JH31	M21137	GAG	Komiyama, N	ARHR 5:411-419 (1989)
B.JP.JH32	M21138	ENV VPU	Komiyama, N	ARHR 5:411-419 (1989)
B.JP.PT1-01	AB034578	VPU	Yamada, T	Arch Virol 145(5):1021-1027 (2000)
B.JP.PT1-4	AB034517	VPR	Yamada, T	Arch Virol 145(5):1021-1027 (2000)
B.JP.PT7-6	AB034474	VIF	Yamada, T	Arch Virol 145(5):1021-1027 (2000)

Table 1: cont.

Name	Accession	Region	Author	Reference
B.JP.nef<7>-a	AB034272	NEF	Yamada, T	Arch Virol 145(5):1021-1027 (2000)
B.KR.CSR9412d	AF238268	NEF	Cho, YK	Unpublished
B.KR.WK	AF224507	ENV GAG NEF POL REV TAT VIF VPR VPU	Cho, YK	Unpublished
B.NL.3202A21	U34604	ENV GAG NEF POL REV TAT VIF VPR VPU	Guillon, C	ARHR 11(12):1537-1541 (1995)
B.SE.AF047085	AF047085	NEF	Visco Comandini, U	J Hum Virol 1(5):320-327 (1998)
B.TH.28-19	U48917	NEF	Artenstein, AW	ARHR 12:557-560 (1996)
B.TH.93TH067	U39258	ENV	Penny, MA	ARHR 12(8):741-747 (1996)
B.TH.AF082839	AF082839	NEF	Vallejo, A	AIDS 13(4):532-534 (1999)
B.TT.QZ4589	U32396	ENV	Blattner, W	Unpublished (1995)
B.TW.TWB101	AF220464	VPU	Lee, CN	J Clin Microbiol 38(7):2468-2474 (2000)
B.TW.TWCYS	AF086817	ENV GAG NEF POL REV TAT VIF VPR VPU	Huang, LM	Unpublished
B.UA.UKR1216	AF193278	ENV REV VPU	Liitsola, K	ARHR 16(11):1047-1053 (2000)
B.US.1-2	U41181	VIF	Sova, P	J Virol 69(4):2557-2564 (1995)
B.US.85WCIPR54	U69584	GAG	Fang, G	J AIDS 12(4): 352-357 (1996)
B.US.AD8	AF004394	GAG	Theodore, TS	ARHR 12(3): 191-194 (1996)
B.US.AF019528	AF019528	VIF	Yedavalli, VR	J Virol 72(2):1092-1102 (1998)
B.US.BC	L02317	VIF	Ghosh, SK	Virology 194, 858-864 (1993)
B.US.DH123	AF069140	ENV GAG	Shibata, R	J Virol 69(7):4453-4462 (1995)
B.US.JRCSF	M38429	ENV GAG NEF POL REV TAT VIF VPR VPU	O'Brien, WA	Nature 348:69-73 (1990)
B.US.JRFL	U63632	ENV GAG NEF POL REV TAT VIF	O'Brien, WA	Nature 348:69-73 (1990)
B.US.LM1	U16909	NEF	Huang, Y	J Virol 69(1):93-100 (1995)
B.US.MNCG	M17449	ENV GAG NEF POL REV TAT VIF VPR VPU	Gurgo, C	Virology 164(2):531-536 (1988)
B.US.NC7	AF049495	NEF	Mwaengo, DM	J Virol 72(11):8976-8987 (1998)
B.US.NY5CG	M38431	GAG	Willey, RL	PNAS USA 83(14):5038-5042 (1986)
B.US.RF	M17451	ENV GAG NEF POL REV TAT VIF VPR VPU	Starcich, BR	Cell 45(5):637-648 (1986)
B.US.SC	M17450	REV	Gurgo, C	Virology 164(2):531-536 (1988)
B.US.SF2	K02007	ENV GAG NEF POL REV VIF VPR VPU	van Beveren, CP	RNA tumor viruses, 2nd edition, Vol 2: 1124-1141; Cold Spring Harbor Laboratory (1985)

Table 1: cont.

Name	Accession	Region	Author	Reference
B.US.WC001	AF003887	REV TAT VIF VPR VPU	Fang, G	J AIDS 12(4):352-357 (1996)
C.BI.BU910112	U39233	ENV	Penny, MA	ARHR 12(8):741-747 (1996)
C.BR.92BR025	U52953	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 70(3):1651-1667 (1996)
C.BW.96BW01B03	AF110959	POL	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW01B21	AF110960	NEF REV VIF VPR	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW01B22	AF110961	GAG TAT	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW0402	AF110962	ENV GAG NEF POL REV TAT VIF VPR VPU	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW0502	AF110967	GAG NEF POL	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW0504	AF110968	REV TAT VIF VPR	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW1104	AF110969	GAG NEF POL REV VIF	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW1210	AF110972	GAG REV	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW15B03	AF110973	GAG REV	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW1626	AF110978	GAG	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW16B01	AF110976	REV	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.BW.96BW17A09	AF110979	GAG REV	Novitsky, VA	J Virol 73(5):4427-4432 (1999)
C.CN.AF268277	AF268277	ENV	Chen, Z	J Virol 74(14):6501-6510 (2000)
C.DJ.DJ259A	L22940	REV TAT VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
C.DJ.DJ373A	L23065	ENV REV TAT	Louwagie, J	J Virol 69(1):263-271 (1995)
C.ET.ETH2220	U46016	ENV GAG NEF POL REV TAT VIF VPR VPU	Salminen, MO	ARHR 12(14):1329-1339 (1996)
C.FR.HIV232980	AJ232980	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
C.FR.HIV232996	AJ232996	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
C.IN.93IN101	AB023804	ENV GAG NEF REV TAT VPR VPU	Mochizuki, N	ARHR 15(14):1321-1324 (1999)
C.IN.93IN904	AF067157	GAG NEF	Lole, KS	J Virol 73(1):152-160 (1999)
C.IN.93IN905	AF067158	GAG	Lole, KS	J Virol 73(1):152-160 (1999)
C.IN.93IN999	AF067154	GAG NEF POL	Lole, KS	J Virol 73(1):152-160 (1999)
C.IN.94IN11246	AF067159	GAG NEF POL REV TAT VIF VPR	Lole, KS	J Virol 73(1):152-160 (1999)
C.IN.95IN21068	AF067155	GAG NEF POL REV TAT VIF VPR	Lole, KS	J Virol 73(1):152-160 (1999)
C.IN.AF209990	AF209990	GAG	Gupta, S	Protein Epr Purif 21(7):378-385 (2001)
C.IN.HIVY15117	Y15117	NEF	Ahmad, KM	ARHR 14(16):1491-1493 (1998)

Table 1: cont.

Name	Accession	Region	Author	Reference
C.IN.HIVY17884	Y17884	NEF	Ahmad, KM	ARHR 14(16):1491-1493 (1998)
C.IN.HIVY17891	Y17891	NEF	Ahmad, KM	ARHR 14(16):1491-1493 (1998)
C.IN.HIVY17892	Y17892	NEF	Ahmad, KM	ARHR 14(16):1491-1493 (1998)
C.SN.SE364A	L22944	VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
C.SO.SO145A	L22946	ENV REV VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
C.TW.TWC2	AF220473	VPU	Lee, CN	J Clin Microbiol 38(7):2468-2474 (2000)
C.UG.UG268A2	L22948	ENV REV VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
D.CD.84ZR085	U88822	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 72(7):5680-5698 (1998)
D.CD.ELI	K03454	ENV GAG NEF POL REV TAT VIF VPR VPU	Alizon, M	Cell 46(1):63-74 (1986)
D.CD.JY1	J03653	ENV	Yourno, J	ARHR 4:165-173 (1988)
D.CD.NDK	M27323	ENV GAG NEF POL REV TAT VIF VPR VPU	Spire, B	Gene 81:275-284 (1989)
D.CD.Z2Z6	M22639	GAG POL REV TAT VIF VPR	Srinivasan, A	Gene 52:71-82 (1987)
D.CI.CI13	AJ277820	ENV	Beirnaert, E	Virology 281(2):305-314 (2001)
D.JP.PT14-4	AB034541	VPR	Yamada, T	Arch Virol 145(5):1021-1027 (2000)
D.SN.SE365A2	L22945	ENV REV TAT VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
D.TZ.87TZ4622	U65075	ENV	Robbins, KE	ARHR 12(14):1389-1391 (1996)
D.TZ.TZ005	U12406	VPU	Siwka, W	ARHR 10(12):1753-1754 (1994)
D.UG.92UG024-D	U08805	ENV	WHO Global Programme	ARHR 10(11):1327-1343 (1994)
D.UG.94UG1141	U88824	GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 72(7):5680-5698 (1998)
D.UG.U18-0	X91355	VIF	Wieland, U	J Gen Virol 78:393-400 (1997)
D.UG.U25-6	X91361	VIF	Wieland, U	J Gen Virol 78:393-400 (1997)
D.UG.U36-0	X91363	VIF	Wieland, U	J Gen Virol 78:393-400 (1997)
D.UG.UG266A2	L22947	VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
D.UG.UG269A	L22949	REV	Louwagie, J	J Virol 69(1):263-271 (1995)
D.UG.UG274A2	L22950	REV TAT	Louwagie, J	J Virol 69(1):263-271 (1995)
D.UG.WHO15-474	U36886	ENV	Douglas, NW	AIDS 10(1):39-46 (1996)
D.ZR.AF233690	AF233690	VIF	Kuhn, J	Unpublished
F1.BE.VI850	AF077336	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	Virology 269(1):95-104 (2000)

Table 1: cont.

Name	Accession	Region	Author	Reference
F1.BR.93BR020-1	AF005494	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 72(7):5680-5698 (1998)
F1.BR.BZ126	L22082	ENV REV TAT VPU	Louwagie, J	ARHR 10(5):561-567 (1994)
F1.BR.BZ162	L11751	GAG	Louwagie, JJ	AIDS 7:769-780 (1993)
F1.BR.BZ163	L22085	REV TAT VPU	Louwagie, J	ARHR 10(5):561-567 (1994)
F1.CD.VI174	L11782	GAG	Louwagie, JJ	AIDS 7:769-780 (1993)
F1.DE.AF200475	AF200475	VIF	Kuhn, J	Unpublished
F1.FI.FIN9363	AF075703	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	Unpublished
F1.FR.MP411	AJ249238	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	ARHR 16(2):139-151(2000)
F1.RW.VI69	L11796	GAG	Louwagie, JJ	AIDS 7:769-780 (1993)
F2.CM.CA20	AJ277824	ENV	Nyambi, PN	J Virol 70(9):6235-6243 (1996)
F2.CM.HIM277819	AJ277819	ENV	Beirnaert, E	Virol 281(2):305-314 (2001)
F2.CM.HIV232985	AJ232985	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
F2.CM.MP255	AJ249236	ENV GAG POL REV TAT VIF VPR VPU	Peeters, M	ARHR 16(2):139-151(2000)
F2.CM.MP257	AJ232986	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
F2.CM.MP257	AJ249237	ENV GAG POL REV TAT VIF VPR VPU	Peeters, M	ARHR 16(2):139-151(2000)
G.BE.DRCBL	AF084936	ENV GAG NEF POL REV TAT VIF VPR VPU	Debyser, Z	ARHR 14(5):453-459 (1998)
G.CG.CNG30	AF056186	VIF VPR VPU	Harada, Y	Unpublished
G.FI.HH8793-1-1	AF061640	GAG NEF POL REV VIF VPR	Salminen, MO	ARHR 8(9):1733-1742 (1992)
G.FI.HH8793-12-1	AF061641	ENV TAT VPU	Salminen, MO	ARHR 8(9):1733-1742 (1992)
G.GA.LBV217	U09664	ENV	Janssens, W	ARHR 10:877-879 (1994)
G.ML.HIV232990	AJ232990	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
G.NG.92NG083	U88826	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 72(7):5680-5698 (1998)
G.NG.IKCSW22	AJ232991	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
G.NG.MACSW39	AJ232992	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
G.NG.NG1937	AF069937	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)
G.NG.NG1939	AF069935	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)
G.SE.SE6165	AF061642	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	Virology 247(1):22-31 (1998)

Table 1: cont.

Name	Accession	Region	Author	Reference
G.TW.TWG1	AF220486	VPU	Lee, CN	J Clin Microbiol 38(7):2468-2474 (2000)
H.BE.VI991	AF190127	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	AIDS 14(11):1533-1543 (2000)
H.BE.VI997	AF190128	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	AIDS 14(11):1533-1543 (2000)
H.CD.HIV232994	AJ232994	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
H.CD.HIV232995	AJ232995	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
H.CF.90CF056	AF005496	ENV GAG NEF POL REV TAT VIF VPR VPU	Murphy, E	ARHR 9(10):997-1006 (1993)
J.SE.SE7022	AF082395	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	ARHR 15(3):293-297 (1999)
J.SE.SE7887	AF082394	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	ARHR 15(3):293-297 (1999)
K.BE.VI325	L11789	GAG	Louwagie, JJ	AIDS 7:769-780 (1993)
K.CD.EQTB11C	AJ249235	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	ARHR 16(2):139-151 (2000)
K.CM.MP535	AJ249239	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	ARHR 16(2):139-151 (2000)
N.CM.YBF106	AJ271370	ENV GAG NEF POL REV TAT VIF VPR	Souquiere, S	Unpublished
N.CM.YBF30	AJ006022	ENV GAG NEF POL REV TAT VIF VPR VPU	Simon, F	Nature Med 4(9):1032-1037 (1998)
O.CM.ANT70	L20587	ENV GAG NEF POL REV TAT VIF VPR VPU	Vanden Haesevelde, M	J Virol 68(3):1586-1596 (1994)
O.CM.CM4974	AF009033	ENV	Korber, BT	Unpublished
O.CM.HIV1CA9EN	X96522	ENV	Janssens, W	AIDS 13:41-48 (1999)
O.CM.MVP5180	L20571	ENV GAG NEF POL REV TAT VIF VPR VPU	Gurtler, LG	J Virol 68:1581-1585 (1994)
O.FR.HIVY16019	Y16019	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16020	Y16020	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16021	Y16021	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16022	Y16022	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16023	Y16023	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16024	Y16024	VIF VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16025	Y16025	VPR	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16026	Y16026	VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)

Table 1: cont.

Name	Accession	Region	Author	Reference
O.FR.HIVY16027	Y16027	VPR	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16028	Y16028	VPR	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16029	Y16029	VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16030	Y16030	VPR	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.FR.HIVY16031	Y16031	VPR VPU	Bibollet-Ruche, F	ARHR 14(11):951-961 (1998)
O.GA.VI686	X96526	ENV	Delaporte, E	AIDS 10(8):903-910 (1996)
O.GQ.193HA	U82990	ENV	Hunt, JC	ARHR 13(12):995-1005 (1997)
O.SN.MP1299	AJ302646	GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	Unpublished (2000)
O.SN.MP1300	AJ302647	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	Unpublished (2000)
CPZ.CD.CPZANT	U42720	ENV GAG NEF POL REV TAT VIF VPR VPU	Vanden Haesevelde, MM	Virology 221(2):346-350 (1996)
CPZ.CM.CAM3	AF115393	ENV GAG NEF POL REV TAT VIF VPR VPU	Corbet, S	J Virol 74:529-534 (2000)
CPZ.CM.CAM5	AJ271369	ENV GAG NEF POL REV TAT VIF VPR VPU	Souquiere, S	Unpublished
CPZ.GA.CPZGAB	X52154	ENV GAG NEF POL REV TAT VIF VPR VPU	Huet, T	Nature 345(6273):356-359 (1990)
CPZ.US.CPZUS	AF103818	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	Nature 397(6718):436-441 (1999)
01_AE.CF.90CF11697	AF197340	ENV GAG NEF POL REV TAT VIF VPR VPU	Anderson, JP	J Virol 74(22):10752-10765 (2000)
01_AE.CF.90CF402	U51188	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	J Virol 70(10):7013-7029 (1996)
01_AE.CF.90CF4071	AF197341	ENV GAG NEF POL REV TAT VIF VPR VPU	Anderson, JP	J Virol 74(22):10752-10765 (2000)
01_AE.CM.CA10	AJ277818	ENV	Beirnaert, E	Virology 281(2):305-314 (2001)
01_AE.DE.K08DE	AF215859	VIF	Kuhn, J	Unpublished
01_AE.FR.HIV232982	AJ232982	NEF	Jubier-Maurin, V	ARHR 15(1):23-32 (1999)
01_AE.TH.93TH057	AF197338	GAG NEF POL REV TAT VIF VPU	Anderson, JP	J Virol 74(22):10752-10765 (2000)
01_AE.TH.93TH065	AF197339	ENV GAG POL REV TAT	Anderson, JP	J Virol 74(22):10752-10765 (2000)
01_AE.TH.93TH253	U51189	POL REV TAT	Gao, F	J Virol 70(10):7013-7029 (1996)
01_AE.TH.93TH902	AF170549	GAG POL VPR	Chang, SY	ARHR 15(17):1591-1596 (1999)
01_AE.TH.94TH702	AF170545	POL VPR	Chang, SY	ARHR 15(17):1591-1596 (1999)

HIV-1/SIVcpz proteins

Table 1: cont.

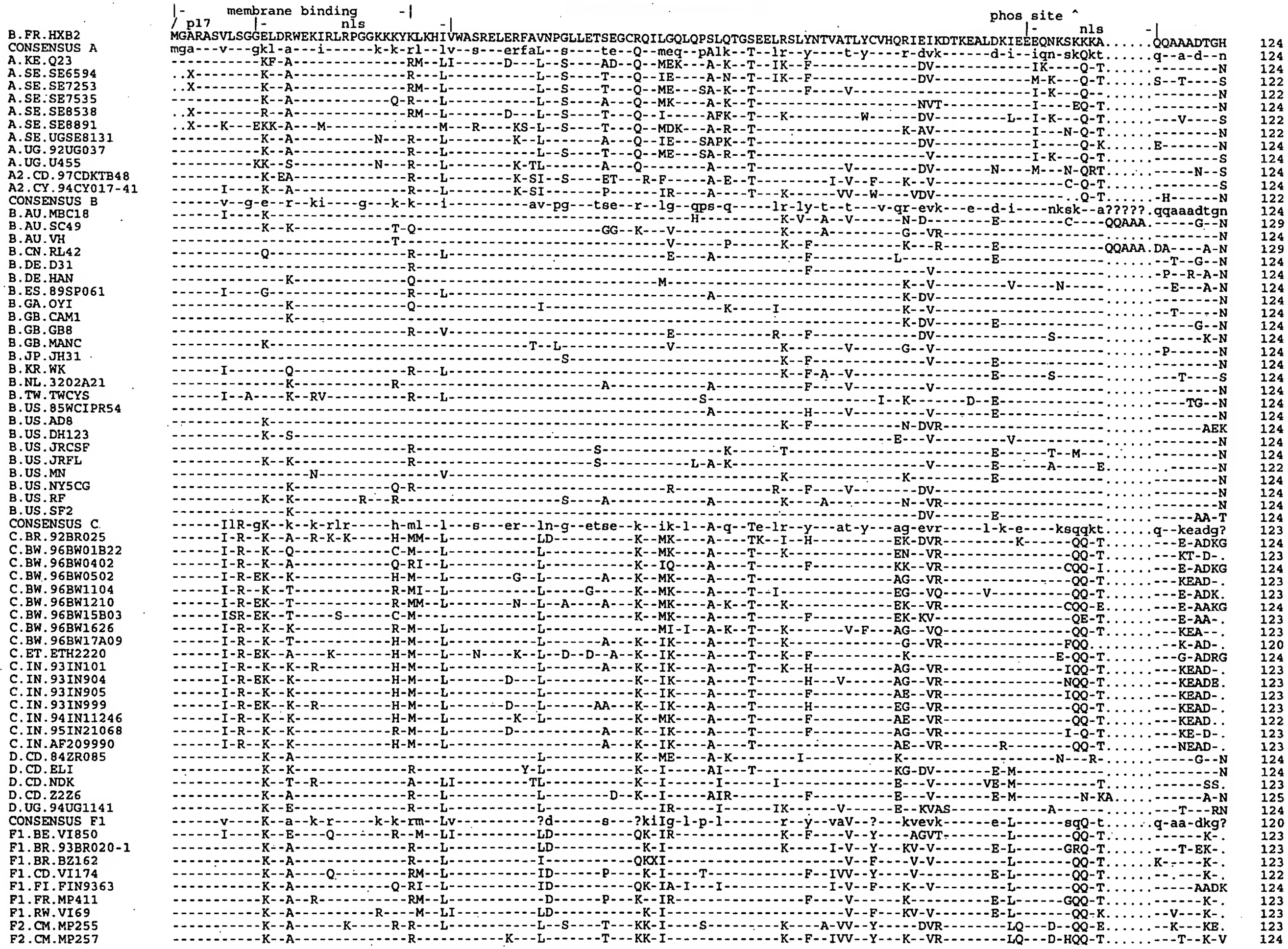
Name	Accession	Region	Author	Reference
01_AE.TH.94TH7091	AF170546	GAG	Chang, SY	ARHR 15(17):1591-1596 (1999)
01_AE.TH.95TNIH047	AB032741	GAG POL REV TAT VIF VPR VPU	Auwanit, W	Unpublished (1999)
01_AE.TH.98-4	U48934	NEF	Artenstein, AW	ARHR 12:557-560 (1996)
01_AE.TH.CM240	U54771	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	J Virol 70(9):5935-5943 (1996)
01_AE.TH.KH03	U48264	ENV	McCutchan, FE	J Virol 70(6):3331-3338 (1996)
01_AE.TH.TH022	AB032740	ENV GAG NEF POL REV TAT VIF VPR VPU	Auwanit, W	Unpublished (1999)
01_AE.TW.TWE13	AF220479	VPU	Lee, CN	J Clin Microbiol 38(7):2468-2474 (2000)
01_AE.TW.TWE6	AF220478	VPU	Lee, CN	J Clin Microbiol 38(7):2468-2474 (2000)
02_AG.CM.MP807	AJ286133	NEF POL REV TAT VIF VPR VPU	Montavon, C	J AIDS 23(5):363-374 (2000)
02_AG.DJ.DJ258A	L22939	VPU	Louwagie, J	J Virol 69(1):263-271 (1995)
02_AG.FR.DJ263	AF063223	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	Virology 247(1):22-31 (1998)
02_AG.FR.DJ264	AF063224	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	Virology 247(1):22-31 (1998)
02_AG.GH.G829	AF184155	GAG NEF POL REV TAT VIF VPR VPU	Candotti, D	J Med Virol 62(1):1-8 (2000)
02_AG.NG.IBNG	L39106	ENV GAG NEF POL REV TAT VIF VPR VPU	Howard, TM	ARHR 10(12):1755-1757 (1994)
02_AG.NG.NG1921	AF069941	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)
02_AG.SE.SE7812	AF107770	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	Unpublished
02_AG.SN.MP1211	AJ251056	ENV GAG NEF POL REV TAT VIF VPR VPU	Toure-Kane, C	ARHR 16(6):603-609 (2000)
02_AG.SN.MP1213	AJ251057	NEF POL REV TAT VIF VPR	Toure-Kane, C	ARHR 16(6):603-609 (2000)
03_AB.RU.KAL153-2	AF193276	ENV GAG NEF POL REV TAT VIF VPR VPU	Liitsola, K	AIDS 12(14):1907-1919 (1998)
03_AB.RU.KAL68-1	AF082485	ENV	Liitsola, K	AIDS 12(14):1907-1919 (1998)
03_AB.RU.RU98001	AF193277	ENV GAG NEF POL REV TAT VIF VPR VPU	Liitsola, K	ARHR 16(11):1047-1053 (2000)
04_cpx.CY.94CY032-3	AF049337	ENV GAG NEF POL REV	Gao, F	J Virol 72(12):10234-10241 (1998)

Table 1: cont.

Name	Accession	Region	Author	Reference
		TAT VIF VPR VPU		
04_cpx.GR.97PVCH	AF119820	ENV GAG NEF POL REV TAT VIF VPR VPU	Nasioulas, G	ARHR 15(8):745-758 (1999)
04_cpx.GR.97PVMY	AF119819	ENV GAG NEF POL REV TAT VIF VPR VPU	Nasioulas, G	ARHR 15(8):745-758 (1999)
05_DF.BE.VI1310	AF193253	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	Virology 269(1):95-104 (2000)
05_DF.BE.VI961	AF076998	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	Virology 269(1):95-104 (2000)
06_cpx.AU.BFP90	AF064699	ENV GAG NEF POL REV TAT VIF VPR VPU	Oelrichs, RB	ARHR 14(16):1495-1500 (1998)
06_cpx.ML.95ML127	AJ288982	ENV GAG NEF POL REV TAT VIF VPR VPU	Montavon, C	ARHR 15(18):1707-1712 (1999)
06_cpx.ML.95ML84	AJ245481	ENV GAG NEF POL REV TAT VIF VPR VPU	Montavon, C	ARHR 15(18):1707-1712 (1999)
06_cpx.NG.NG3670a	AF069934	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)
06_cpx.SN.97SE1078	AJ288981	ENV GAG NEF POL REV TAT VIF VPR VPU	Montavon, C	ARHR 15(18):1707-1712 (1999)
10_CD.BFL061	AF289548	ENV GAG NEF POL REV TAT VIF VPR VPU	Koulinska, IN	ARHR 20(5):423-431(2001)
10_CD.BFL071	AF289549	ENV GAG NEF POL REV TAT VIF VPR VPU	Koulinska, IN	ARHR 20(5):423-431(2001)
10_CD.BFL110	AF289550	ENV GAG NEF POL REV TAT VIF VPR VPU	Koulinska, IN	ARHR 20(5):423-431(2001)
11_cpx.CM.CA1	AJ277823	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)
11_cpx.CM.MP818	AJ291718	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	Unpublished (2000)
11_cpx.FR.MP1298	AJ291719	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	Unpublished (2000)
11_cpx.FR.MP1307	AJ291720	ENV GAG NEF POL REV TAT VIF VPR VPU	Peeters, M	Unpublished (2000)
11_cpx.GR.GR17	AF179368	ENV GAG NEF POL REV TAT VIF VPR VPU	Paraskevis, D	ARHR 16(9):845-855 (2000)
11_cpx.NG.NG3670b	AF069945	ENV	McCutchan, FE	Virology 254(2):226-234 (1999)

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins



```
                    |-     membrane binding     -|
                    / p17     |-        nls       -|                                                
```

```
                                p17 \/ p24                                                
```

```
                                                
```

HIV-1/SIVcpz proteins

major homology region (MHR)

p24 \/ p2

```
B.FR.HXB2           LQEQIGWMTN...NPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQ 369
CONSENSUS A         p---ig---g...---i---D--k--i----------V-----k----------F-tl-----t-e--g---e---v--------s--ra--aG-------t--------g---------- 369
A.KE.Q23            P--------G...-------D------------------V-----K----------F--F-----T-D------D-----------A--R----G---------------------------- 369
A.SE.SE6594         P--------S...-------D------------------V-----K----------F-V-----T----G-------I--------S--R---TG---------------------------- 367
A.SE.SE7253         P--------G...-------D------------------V-----K----------F----------D----------I--------S--R---AG-------S--------S---------- 367
A.SE.SE7535         P---LQ---S...-------D------------------V----------------F-A------T----G----------------S------TG---------------------------- 368
A.SE.SE8538         P--------S...-------D------------------V-----K----------F--------T----G----------I--------S--R----G---------------------------- 367
A.SE.SE8891         -----A---S...---V---D------------------V-----K----------F-A------T----G---D----------S--R---SG---------------------------- 367
A.SE.UGSE8131       P----A---G...-------D-----M----------V-----K----------F--------T----G-------I--------S--R---AG-------S--------S---------- 369
A.UG.92UG037        V--------G...-------D--R---------------V----------------F--------T-D------------------S--R----G---------------------------- 369
A.UG.U455           ---------S...-------------------------V-----K----------F--------T----G---------------S--R----G---------------------------- 365
A2.CD.97CDKTB48     ---------S...D------------------------V----------------F--------T--------D----------S--R----G-S-----------S----------H 375
A2.CY.94CY017-41    -----g---n...--pi----------l----------ts----r-----p---------k---a---s-e---wm---l----a----k--------a----------g----v------q 369
CONSENSUS B         ---------G...--A---------------------I-----K----------------------T---------------------------------------S---------- 374
B.AU.MBC18          ---------S...---V----------------------V-------------------------------------D-------------------------------S----I------- 369
B.AU.SC49           ----------...-------------------------------K----------R-------------------------R---------------------------------------- 375
B.AU.VH             ----------...-----------------------------C-------------------------------D-------------------------------S----I------- 369
B.CN.RL42           ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.DE.D31            -----A----...-------------------------------K----------------------T---------------------------------------------------- 369
B.DE.HAN            ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.ES.89SP061        ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.GA.OYI            ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.GB.CAM1           ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.GB.GB8            ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.GB.MANC           ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.JP.JH31           ---------S...-------------------------------K----------------------T---------------------------------------------------- 369
B.KR.WK             ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.NL.3202A21        ---------H...-------------------------------K----------------------T---------------------------------------------------- 369
B.TW.TWCYS          ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.85WCIPR54      ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.AD8            ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.DH123          ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.JRCSF          ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.JRFL           ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.MN             ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.NY5CG          ---------H...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.RF             ----------...-------------------------------K----------------------T---------------------------------------------------- 369
B.US.SF2            ----------...-------------------------------K----------------------T---------------------------------------------------- 369
CONSENSUS C         -----a---s...---v---d------il----------V----ik-----p--------Fk------at-d------d----------kt--r-l--g-sl----------s---rv-----sq 371
C.BR.92BR025        -----T----...---V---D------------------V-----K----------F--------T-D------D----------R----G-S-----------------------K 366
C.BW.96BW01B22      -----A---A...---V---D------------------V-----K----------F--------T-D------D----------R----G-SI----------S----------G- 367
C.BW.96BW0402       -----A---S...-------D------------------V-----K----------F-------ST--------D----------R----G-S-----------S---------- 366
C.BW.96BW0502       -----A---S...---V---D------------------V----------------F--------T-D----------------R----G---------------------------- 367
C.BW.96BW1104       -----A---S...-------D------------------V-----K-----S----F-------S---------D----------R--------G-S----------S----I------- 366
C.BW.96BW1210       -----N---S...-------D------------------V-----K----------F--------T-D------D----------R----G-S----------S----------H 366
C.BW.96BW15B03      -----A---S...-------D------------------V----------------F--------T-D------D----------P--R----G---------------------------- 367
C.BW.96BW1626       -----A---S...---V----------------------V----------------F--------T-D------D----------R-S--G-S---------------------------- 366
C.BW.96BW17A09      -----A----...-------D------TM----------V-----K----------F--------T-D------D----------R----G----------S----------H 366
C.ET.ETH2220        -----A---G...---V---D------------------V-----K----------F--------T-D------D----------R----G-S----------A---------- 363
C.IN.93IN101        -----A---G...---V---D------------------V----------------F--------T-D------D----------R----G-S----------S---------- 367
C.IN.93IN904        -----A---G...---V---D------------------V-----K----------F--------T-D------D----------R----G-S----------S---------- 366
C.IN.93IN905        -----A---G...---V----------------------V-----K----------F--------T-D------D----------R------S---------------------- 366
C.IN.93IN999        -----A---G...---V---D------------------V-----K----------FR-------T-D------D----------R----G-S---------------------- 366
C.IN.94IN11246      -----A---S...---V----------------------V----TK----------F--------T-D------D----------R----G-S----------S---K-------- 366
C.IN.95IN21068      -----A----...---V---D------------------V----------------F--------T-D----------------R----G-S----------S---------- 363
C.IN.AF209990       -----A---G...---V---D------------------V-----K----------F--------T-D------D----------R----G-S----------S---------- 366
D.CD.84ZR085        ---------S...-------------------------V----------------------------------------------G----------S---------- 366
D.CD.ELI            -----A---S...----------------V---------V----------------------------D----------------Q----------S---------- 369
D.CD.NDK            -----A---S...-------------------------V----------------------------D----------------Q---------------------- 369
D.CD.Z2Z6           -----A---S...-------------------------V----------------------G----------------------Q----------S---------- 366
D.UG.94UG1141       ---------S...-------------------------V----------------------------D----------------------------S---------- 370
CONSENSUS F1        -----Q---s...---v---di----------------Vs----r----------F-t----q-t-e--g---dt--v--a-----t---al-pG----------q-----g----v---a--- 369
F1.BE.VI850         -----Q---G...---V---D------------------V-----K----------F-V--------D--G---D----------TG----------S---------- 361
F1.BR.93BR020-1     -----Q---G...---V----M-----------------VG----------F--------T----G---D----------G----------S---------- 365
F1.BR.BZ162         -----Q---S...---V----------------------V----------F--------T----G---D----------G---------------------------- 365
F1.CD.VI174         -----Q----...---V----------------------V----------F--------T----G---D----------G---------------------------- 365
F1.FI.FIN9363       -----Q---S...---V---D------------------V----------F-A------T----G---D----------I---G--IG----------R----------I------- 364
F1.FR.MP411         -----Q---S...---V---D------------------V----------F----------------S-----S----------G---------------------------- 367
F1.RW.VI69          -----Q---S...-------D------------------V----------F------E-T----G---D---I----------M--G----------------------T---- 365
F2.CM.MP255         -----T---G...---V----------------------V-----K----------F--------T----------S----------G----------I------K 365
F2.CM.MP257         -----A---S...---V----------------------V-----K----------F--------T----G----------------G----------S----I------K 369
```

```
                                                major homology region (MHR)
                                                          |-                -|                                                
```

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

```
                                                vpr binding
                                                |-    -|        \ p6 terminus
B.FR.HXB2          ............TPPQKQEP......IDKELY.PLTSLRSLFGNDPSSQ$          500
CONSENSUS A        ............aS-pk--q.....?kdreqnp-lv--K------lL---          498
A.KE.Q23           ............VS-L---Q......K-R-QAQ--V--K------LL---          498
A.SE.SE6594        ............AS-PR--Q..........TP-SV--K-------L---          491
A.SE.SE7253        ............AS-P---Q......K-R-QNS-SV--K------LL---          497
A.SE.SE7535        ............AS-P---Q........-GQDP--V--K------LL---          495
A.SE.SE8538        ............-S-P---Q......K---QVP--V--K-------L---          496
A.SE.SE8891        ............-S-P---K.....KDRESEAP--I--K-------L---          497
A.SE.UGSE8131      ............AS-P---Q.........NNP-SV--K------LL---          492
A.UG.92UG037       ............VS-P---Q......N-RDQNP-SV--K------LL---          499
A.UG.U455          ............-S-A---L......K-R-QT.--V--K-------L---          493
A2.CD.97CDKTB48    ............-SSL---N........R-PST-AI--K------LL---          497
A2.CY.94CY017-41   ............-SSL---L......ETR-P-N-AI--K-------LL--          496
CONSENSUS B        ....????????t-s----p??????idkel-.-la--rs---n--ssq-          500
B.AU.MBC18         ..............--S-------...IDR-SS--.--A--------------          517
B.AU.SC49          ....TPPQKQEPL-S----T.....---D--.--A--K--------L--          508
B.AU.VH            ..............--S-----.....------.--A--------------          506
B.CN.RL42          ..............--S-----.....------.--A--K-----------          505
B.DE.D31           ..............--F-----.....------.--A--------------          500
B.DE.HAN           ............A-S-----.....------.--A--K----S------          499
B.ES.89SP061       ..............--S----L.....-------.-SA--K-------L---          503
B.GA.OYI           ..............--------.....---G--.----------------          499
B.GB.CAM1          ..............--S-------.....------.--A--------------          500
B.GB.GB8           ..............--S-------.....-N--P-.--A--------------          507
B.GB.MANC          ..............--A-----.....------.--A---P-----------          500
B.JP.JH31          ..............--S-------.....R------.--A----------P-          500
B.KR.WK            ..............--S-----.....------.--A--------------          500
B.NL.3202A21       ..............--S-------.....R------.--A--------------          500
B.TW.TWCYS         ..............--S-------.....---D--.--A--E-----------          500
B.US.85WCIPR54     ..............--S-------.....----Q-.--A--------------          500
B.US.AD8           ..............--S-----.....------.-----K-----------          500
B.US.DH123         ..............--S-----.....-------.--A--K-------...-          495
B.US.JRCSF         ..............--S----QKQEP..------.----------------          504
B.US.JRFL          ..............--S----------.----M-.----------------          498
B.US.MN            ..............--Y----KKQET..---D--.--A--K-------L---          506
B.US.NY5CG         ..............--S-------.....------.--A-------S------          500
B.US.RF            ..............--S----K.....------.--A--K-----------          501
B.US.SF2           ..............--S-------.....------.----------------          502
CONSENSUS C        ............?-apk-ep.....?kdre???p-tslk----sdpLsq-          491
C.BR.92BR025       ..............--SR---T.....------..-----K----S--L-T-          496
C.BW.96BW01B22     ..............-AP---------.K-R-...----------S--L---          492
C.BW.96BW0402      ..............-VP---------.K-R-...-----K----S--L---          492
C.BW.96BW0502      ..............-AP---------.KDREPYRE---A------SG-L---          502
C.BW.96BW1104      ..............-AP---T.....K-R-...--I--K----S--L---          490
C.BW.96BW1210      ..............-A----------.K-R-..P--A--K-------L---          494
C.BW.96BW15B03     ..............-AP---------.K-R-...--I--K----S--L---          492
C.BW.96BW1626      ..............-AP---------.K-R-...-------------L---          492
C.BW.96BW17A09     ..............-AP---------.K-R-...----FK----S--L---          490
C.ET.ETH2220       ..............-SP---L.....K-R-...A----K------HLL--          504
C.IN.93IN101       ..............-AP---------.K-R-...-----K----S-LL---          491
C.IN.93IN904       ..............-AP---------.K-R-...----------S--L---          491
C.IN.93IN905       ..............-AP---------.K-R-...----------S--L---          491
C.IN.93IN999       ..............-AL---------.K-R-...-----K----S--L---          500
C.IN.94IN11246     ..............-AP---------.KER-...----------S--L---          488
C.IN.95IN21068     ..............-AP---------.K-R-...----------S--L---          491
C.IN.AF209990      ..............-ALQ-G-.....K-R-...----------S--L---          491
D.CD.84ZR085       ..............-S----Q....KDK-----.--A--K-------L---          502
D.CD.ELI           ..............-S----Q.....K-----.-----K-------L---          500
D.CD.NDK           ..............-S----Q.....K-----.--A--K-----------          497
D.CD.Z2Z6          ..............-S----Q.....K-----.-S-A-K-------LL--          501
D.UG.94UG1141      ..............------Q.....K-----.-----K-------L---          502
CONSENSUS F1       ............?-SPk--Q....??kdeglyP--A--K----n--...?          488
F1.BE.VI850        ..............-SP---Q.....K-G---P--A--K-------...-          491
F1.BR.93BR020-1    ..............-SP---Q.....K-EG--P--A--K-------...-          492
F1.BR.BZ162        ............N-SP---Q.....KEEG--P--A--K-----------          486
F1.CD.VI174        ..............-SP---Q.....KGEG--P--A--K-----------          487
F1.FI.FIN9363      ..............-SPR--Q.....KEEGQ-P--A--K-------...-          493
F1.FR.MP411        ..............-SP---Q....KDEGQG--P--A--K----S--...-          494
F1.RW.VI69         ..............-SP---Q.....K-EG-SP--A--K-----------          486
F2.CM.MP255        ..............-SP---Q.....KGE-QAP--V--K----S--...-          492
F2.CM.MP257        ..............-SP---Q.....K---QVP--I--K----S-Q...-          496
```

```
                                                  vpr binding
                                                  |-    -|       \ p6 terminus
B.FR.HXB2           ............TPPQKQEP......IDKELY.PLTSLRSLFGNDPSSQS$          500
G.BE.DRCBL          ............-SP---Q......KE----.--S--K------Q...-          490
G.FI.HH8793         ............-SP-P--......KE--TH.--A--K----S--...-          497
G.NG.92NG083        ............-SP---.......KE----.-----K----S--...-          497
G.SE.SE6165         ............-SP---M......KE----P.....K----S--...-          495
H.BE.VI991          ............-SPR--L......KEQ-..P-----------Q...-          498
H.BE.VI997          ...........SSP---L......K---..P-FA--K-------L---          499
H.CF.90CF056        ............-SP---Q.....LK---..P--A-------S--LL--          500
J.SE.SE7022         ............-SP---.......K-----.----------S--L---          497
J.SE.SE7887         ............-SP---.......K-----.-----K----S--L---          497
K.BE.VI325          ............-SPR--T......K---QGS-----K-.........          486
K.CD.EQTB11C        ............-SLR--M......K-Q-QGP-----K----S--L---          495
K.CM.MP535          ............-SPR--T......K---QSP-----K-------L---          495
N.CM.YBF106         ..........KS-QGKEMQENQEK..TETS--P-----------LX---          511
N.CM.YBF30          ..........KS-QGKEMQENQER..TENS--P----------------          512
O.CM.ANT70          .............V-GQENQEQKGGPN---.-FA--K----T-...---          498
O.CM.MVP5180        .............V-EQENQSQKGDQE---.-FA--K----T-...--          498
O.SN.MP1299         .............M-GQENQEQKEDQN---.-FA-------T-...--          497
O.SN.MP1300         .............M-GQENQEQKGDQN---.-FA--K----T-...--          497
CONSENSUS 01        ?...........-Sf?kqeq......K-kehpp-lv--K-------l---          496
01_AE.CF.90CF11697  ............-SFP---Q......K---HPS--V--K-------L---          498
01_AE.CF.90CF402    ............-SFP---Q......K--KQPP--V--K-------L---          498
01_AE.CF.90CF4071   ............-SFSR--Q......K-R-HPP--V--K-------L---          498
01_AE.TH.93TH057    ............-SLL---Q......K---HHP--V--K-----------          498
01_AE.TH.93TH065    ............-SLP---Q......K--D.PP--V--K-------L---          496
01_AE.TH.93TH902    ............-SSL---Q......K---YPP--V--K-------L---          498
01_AE.TH.94TH7091   ............-SAXNEDE.....K---YPP--V--K-------L---          498
01_AE.TH.95TNIH047  ............-SSL---Q......K---PPP--I--K-------L---          498
01_AE.TH.CM240      I...........-SLP---Q......K---HPP--V--K-------L---          503
01_AE.TH.TH022      ............-SFL---Q......K---HPP-SV--K-------L---          498
CONSENSUS 02        ............tsspk--p......r-kG--P--?--K-------???-          483
02_AG.FR.DJ263      ............-S-P---A......R-QG--P--A--K--------...-          492
02_AG.FR.DJ264      ............-S-P---.......R-QG--P--A--K--------...-          492
02_AG.GH.G829       ............-SSPQ---......R--G--P-----K--------...-          491
02_AG.NG.IBNG       ............P-SPQ---......R--G--P-----K--------...-          492
02_AG.SE.SE7812     ............-SS----......G--G--P--A--K--------...-          492
02_AG.SN.MP1211     ............-SSP---.......G--G--P-----K-------Y---          495
03_AB.RU.KAL153-2   ............--SL---Q......K-R-QHP-SI--K----D--L---          498
03_AB.RU.RU98001    ............--SL---Q......K-RGQHP-SI--K-------L---          498
04_cpx.CY.94CY032-3 ............-SSL---.......R-----.-----K----S--L---          500
04_cpx.GR.97PVCH    ............-SSP---.......R-----.-----K-----H-L---          500
04_cpx.GR.97PVMY    ............-SSP---Q......R-----.-I---K----S--L-R-          498
05_DF.BE.VI1310     ............ASSP---Q......K-EG--P--A--K-------L---          502
05_DF.BE.VI961      ............--SP---Q......K-EGK-P--A--K-------L---          501
06_cpx.AU.BFP90     ............A-SP---S...KEKEE-G--.--A--K----S--...-          497
06_cpx.ML.95ML127   ............A-SPE-K-.....KE----.-------------...-          497
06_cpx.ML.95ML84    ............A-SL---....KEKE----.--A--K-------...-          499
06_cpx.SN.97SE1078  ............A-SP---.......KE----.--A--K----S--...-          502
10_CD.BFL061        ............--S----Q......K---PH.--A--K-------L---          500
10_CD.BFL071        ............--F----Q.....KDKE---.L-A--K-------L---          500
10_CD.BFL110        ............-S----Q......K-----.--A--K-V-----L---          497
11_cpx.CM.MP818     ............A-SP-------...KEKE----.-----K----S--...-          496
11_cpx.FR.MP1298    ............A-SP---.......KE----.-----K-------L---          497
11_cpx.FR.MP1307    KQEQSFGIGENVA-SP---.......KKE---.-----K-------LL--          513
11_cpx.GR.GR17      ............--SP---.......KE-----.I---K----S--L---          494
CPZ.CM.CAM3         ..............GKRQEEEREKEKEP--P-----K----S----E-          507
CPZ.CD.CPZANT       ..............KRTQKGLKGEE--.P-SY--K----K-...--          522
CPZ.CM.CAM5         ...............MEKKKEEKEP--PQ----K----S---LE-          502
CPZ.GA.CPZGAB       ..............KSQEKKEGESS.--P-.---K----S------          508
CPZ.US.CPZUS        ..............QEKQGKEKEP..FQ----K----S----E-          505
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins


```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins


```
                                                
```

M41L

protease \/ p66, p51 RT

```
B.FR.HXB2          EEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIK  219
H.BE.VI991         -DIN---K-----------------E-VA---F-K-----------------I---M-----L----------T-----------------T---L-----------------I----  217
H.BE.VI997         --IN-L--------------------VA-----K-----------------I-------------------------R-------------T---M-----------------I----  217
H.CF.90CF056       --IN---K-----------------E-VA-----K-----------------I-----------------------------------------T-------------R-------S--I----  218
J.SE.SE7022        --ID---K-----------------NEVP---E-K-------I----------M---L-------------------------I-------------TQ--A-L-E-----R----------  217
J.SE.SE7887        -DID--RK-----------------NEVP---E-K-------I----------M---L-------------------------I-------------TQ--A---E-----RV---------  217
K.CD.EQTB11C       --IN---K--------------------VCM----Q-----------------M-------------------------------------------------------------------  219
K.CM.MP535         --IN---K--------------------V------Q-------------------------------------------------------T-------------------------------  219
N.CM.YBF106        --IQ-E-K-----------------N-T-D-Q-R--V---------------F-------------------------R-------A---E--R----------R----------I----  226
N.CM.YBF30         --LQ-E-K-----------------N-TVD-Q-R--V-------------------------------------------T---E--R----------R----------I----  226
O.CM.ANT70         NNIQ-E-K---------------KE--NVTV--E-REVQ----------------I--GL----------AP----------------SK---E--TA--Q---Q-----R----------I----  215
O.CM.MVP5180       NNIQ-E---T-------------KE-NNVTV-VQ-KEVQ-------------L---I--GL----------AP----------------SR---E--TA--Q---Q-----R----------I----  215
O.SN.MP1299        TNIQ-E-K-T-------------KE-N-VPV--E-REVL----------------I--GL----------AP----------I-----SK---E--TA--Q---Q-----R----------I----  215
O.SN.MP1300        NNIQ-E-K-T-------------KE-N-VPV--E-REVL----------------I--GL----------AP----------I-----SK---E--TA--Q---Q-----R----------I----  215
CONSENSUS 01       -DIN--gk---k-----------r-----l-----Kk----------------M----------iD-v--t----------------T-i-Ke--E----------------v----  219
01_AE.CF.90CF11697 -DIN-------------------------------KR----------------M----------D----T----------$-----$--T---K---E----------------------  217
01_AE.CF.90CF402   -DIN---K-------------------------K-----------------M----------D----T----------------T---K---E----------------------  219
01_AE.CF.90CF4071  -DIN---K---------------------I----K-----------------M----------D----T----------------T---K---E----------------------  219
01_AE.TH.93TH057   -DIN---K-------------------------K-----------------M----------D----T----------------T---K---E----------------------  219
01_AE.TH.93TH065   -DIN---K-------------------------K-----------------M----------XD---T----------------T-T-K---E----------------------  217
01_AE.TH.93TH253   -DIN---K-------------------------K-----------------M----------D----------------T---K---E----------------------  211
01_AE.TH.93TH902   -DIN---K---------------------T----K-----------------M----------D-I--T----------------T---K---E----------------------  219
01_AE.TH.94TH702   -DIN--EK---------------------T----K-----------------M----------D-I--T----------------T---K---E----------------------  219
01_AE.TH.95TNIH047 -DIN---K-------------------------K-----------------M----------D----------------T---K---E----------I----  219
01_AE.TH.CM240     -DIN---K---------------K----------K-----------------M----------D----T----------------T---K---E----------I----  224
01_AE.TH.TH022     -DIN---K---R-----------------------K-----------------M----------D----------------T---KK--E----------------------  219
CONSENSUS 02       -eIn---K--------------v------li----K--i-t----------r-m--q----------v---------k----------i-a-td--t---------k----------v-a--  218
02_AG.CM.MP807     --ID---K-------------------------K-----------------M-------------------------------------TD--L----------------------  219
02_AG.FR.DJ263     --IN---K-------------------------K-----------------M-------------------------------------T----------------------G--  219
02_AG.FR.DJ264     --IN---K-------------------------K-----------------M-------------------------------------TD--A----------------------  219
02_AG.GH.G829      -DIN---K---------------------P-----K-----------------M-------------------------------------TD--M----------I----  219
02_AG.NG.IBNG      -DIN---K-------------------------K-----------------M-------------------------------------TD-----------------------  219
02_AG.SE.SE7812    --IN---K-------------------------K-----------------M-------------------------------------TD----------R----------I----  219
02_AG.SN.MP1211    -KIN---K-------------L----------K--M-S----------K-I----------------L-----------------L-H-TD--A----------------------  219
02_AG.SN.MP1213    -KIN---K-------------L-------M----K----S----------Q-I--P----------------------R-------------I----------------------  219
03_AB.RU.KAL153-2  -DIN---K-------------------------K-----------------M---L----------T-----------------TD--K----------------------  219
03_AB.RU.RU98001   -DIN---K-------------------------K-----------------M---L----------T-----------------ID--K----------------------V--  219
04_cpx.CY.94CY032-3 --IN---K---------------------P-----K-----------------M---L-------------------------T----D----------I----  218
04_cpx.GR.97PVCH   -DIN---K---------------K----------K-------------------------------------------------T----------I----  218
04_cpx.GR.97PVMY   --IN-------------------------T-----K--T---------A------M---L-------------------------R--------Q------V---------I----  218
05_DF.BE.VI1310    -DIN---K-----------------------V----Q-------------------------------------------------T----D--------R----------I----  219
05_DF.BE.VI961     ---N---K-------------------------------V-------------------------------------------T----D--------R----------I----  219
06_cpx.AU.BFP90    -DIN---K---------------K----------K----------I------M-------------------------------T---K----------------------  221
06_cpx.ML.95ML127  -DIN---K-------------------------KR-M----------------M-------------------------------T----------R----------I----  218
06_cpx.ML.95ML84   -DIN---K-------------------------K--M----------------M----------------------I------T----------I----  220
06_cpx.SN.97SE1078 --IN-------------------------P-----KR----------------M-------------------------------T----------I----  223
10_CD.BFL061       ---N---K-----M---------------------Y-------------------------------------------------T---K---------R----------I----  218
10_CD.BFL071       -D-----K----------------E----P-----Y-------------------------------------------------T----D--------R----------I----  219
10_CD.BFL110       ---N---K-------------------------Y-------------------------------------------------T----D--------R----------I----  216
11_cpx.CM.MP818    --IN-------------------K--EE-I---E-K----------------M----------D----Q--A----------------T----------------------------  220
11_cpx.FR.MP1298   --IE---------------------EE-----E-K----------------M----------D-------------------S-------K----------------------  218
11_cpx.FR.MP1307   -DID----------------V-----ED-T---E-K-----------A------MM----------D-----------------------T-----L----------------------  234
11_cpx.GR.GR17     ----------------------K--ED-T---E-K-------I----------M----------VD----------------------T----------------------------  218
CPZ.CD.CPZANT      -GIH-Q-N----T--------S-Q--NKVP-Q-GDRTVLA---L--N--------V-CLL---------KV---------E-----R------SK---E--K---DKL-A-N---R---D------I----  219
CPZ.CM.CAM3        DNVQ-T-K-R-------------K---N-A---E-R-TT----------------I-------------------------R-------A------T---Q---------R----------I----  224
CPZ.CM.CAM5        DNVQ-T-K-R-R-----------K---N-V---E-R-TT----------------I--------------V----------R-------A------T---Q----------------------I----  220
CPZ.GA.CPZGAB      -RIQ-Q-L---------------K-F-NVH---E-R-VV----------------I---L----V----S----------------------SA------T---Q----------------------I----  217
CPZ.US.CPZUS       DNIQIE-T-R---M---------K---HVN---E-R--Q-S--------------I-------------S-K----R--------R-------A------T---Q--------TRV---------I----  221
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

D67N K70R

polymerase motif

```
B.FR.HXB2          KKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLE 349
CONSENSUS A        --d-tk-------------------------p------k------v---------d-s----------t--e---i---y----------a---s---------skn-eii---------------  347
A.KE.Q23           ------------------------------------------------H-E----------T------V-------------------------SK--E---------------  349
A.SE.SE6594        --N--R-----------------------------N-----------H-----------T--A-------X-------------------------SK----I---------------  344
A.SE.SE7253        ------------------------------------------S-----------T------V-------------------------------LK--E-I---------------  349
A.SE.SE7535        ---------------------------------------L---------H-G----------T--A-------------------------------ER--EVI---------------  347
A.SE.SE8538        ----NR-----------------------------------------S----------T-----------------------S-------------SK--E-I---------------  349
A.SE.UGSE8131      --N-DR-----------------------------------------S----------T-------------------------------------SK--E-I---------------  343
A.UG.92UG037       ------------------------------------------S----------T-------------------------A----------SK----------------------  350
A.UG.U455          ----------------------------T-------------------S----------------V---------------S-------------S-H---------------  348
A2.CD.97CDKFE4     ------------------------------------------------------------S----------------S-------------SKY-EM---------------  341
A2.CY.94CY017-41   ---------------------------------A----------H-----------T------V-------------------------SK-TELI---------------  349
CONSENSUS B        ---s------------kr------v-------a----kk-v-v-------------kd----------i-----gi-------------------s---k-----rk---d-v---------------  349
B.AU.MBC18         ---G-----------------------------S-----R-----------------N-------------------------------------R------R---------------  361
B.CN.RL42          ------------------------A-----------------------------K----------V-------------------------C-------------------------  349
B.DE.D31           ------------------------------------------------------------V-------------------------------------------------------  349
B.GA.OYI           ------------------------------------------K-------------------------------------------------------------------------  349
B.GB.CAM1          ------------------------------------------K----------T-------------------------------------------------------------  349
B.KR.WK            ------------------------------------------------------T-------------------------------------------------------------  349
B.NL.3202A21       ---------------------------------------G-A----------K----------V-------------------------C-------------------------  349
B.TW.TWCYS         ------------------------------------------K-------------------------------------------------R---------------------  349
B.US.JRCSF         ----------------R-------------------------K-------------------------------------------------------I---------------  353
B.US.JRFL          ----------------K-----------------R------K-------------------------------------------------------I---------------  349
B.US.MN            ----------------K-----------------------K-------------------------------------------------------------------------  352
B.US.RF            ------------------------------------------KE----------------R-------------------------K----E---------------------  348
B.US.SF2           ------------------------------------------K-------------------------------------------------------------------------  349
CONSENSUS C        ---------------------------------------v---------d-g----------i--e-----------------a---s-mtk------aq--eiv----------s---  345
C.BR.92BR025       -------------------$-----------------------------G-------------------------------S-----T--------A---E-I---------------  347
C.BW.96BW01B03     ------------------------------------------------G-------------------------------------------------AL--E------------P---  345
C.BW.96BW0402      ------------------------------------------------S-------------S-----------------------I-------TK----------------------  345
C.BW.96BW0502      ---------------------------------------M----------G-------------------------------------------------L---E---------------  355
C.BW.96BW1104      ------------------------------------------------S-------------------------------S-------------AK--EL----------------  343
C.ET.ETH2220       ------------------------------------------------G----------T---------------------P------PQ------AP--E---------------  356
C.IN.93IN999       ----------------------------------------------------Y----------R-------------------------A---------A---E---------------  354
C.IN.94IN11246     ----------------------------------------------------Y-------------------------------------------------GR--E-D---------  345
C.IN.95IN21068     ----------------------------------------------------Y-------------------------------------N---R------A---E---------------  344
D.CD.84ZR085       ---------------------------------------I----------I--C-------------------------------------I----------EV----------------  350
D.CD.ELI           ------------------------------------------------------------S-------------------------------------EM----------------  348
D.CD.NDK           ------------------------------------------------------------------------------------------------E-----------------  348
D.CD.Z2Z6          ------------------------------------------K-------------------------------------------------------E-----------------  348
D.UG.94UG1141      ------------------------------------------H----------T-------------------------------------------EMI---------------  350
F1.BE.VI850        ----S--------K-------------------------------------K--K---------V-------------------------C----------MK----------------  349
F1.BR.93BR020-1    ------------------------------------------K-------S----T------V-------------------Y------D---AK--------------------  349
F1.FI.FIN9363      ------------------------------------------K----------V-------------------------C----------TR----------------------  348
F1.FR.MP411        ------------------------------------------KE---------L-------------------------------------AK----------------------  351
F2.CM.MP255        ------------------------------------------KE-------------------------------------C----------AK--E---------------  349
F2.CM.MP257        ---------------------------------R-------KE-------------------------------------I--------E--E---------------  349
G.BE.DRCBL         -----R-------------------P------G---Q-R----------N----------T----------X..-------------------T---E---------------  346
G.FI.HH8793-1-1    ---------------------------------RENQSA---------S----------T-------------------------------IK--EM----------------  348
G.NG.92NG083       ---------------------------------R-------------K-------------------------------------S-TK--EM----------------  348
G.SE.SE6165        ------------------------------------------------------------V-------------------R------AN--EM----------------  346
```

D67N K70R

polymerase motif

```
B.FR.HXB2                KKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLE 349
H.BE.VI991               --N--R-----------------------------------S-----G-------H-----------T-----------------------------------------EVI---------- 347
H.BE.VI997               ----------------------------------------S-------------K-------------------------------------------------E-I---------- 347
H.CF.90CF056             ----------------------------------------S-------------KE------------------------------------------------------------ 348
J.SE.SE7022              ------------------------------------------------------Y-----------------------------------------A---E---EM---------- 348
J.SE.SE7887              ------------------------------------------------------Y-----------------------------------C------K---ER--E---------- 347
K.CD.EQTB11C             -------I-----------P-------------------------------------K-----------------------------------C----------RK---M-L------- 347
K.CM.MP535               ------------------------------------------------------K-----------------V-----------------H----------IK--EM---------- 349
N.CM.YBF106              ---------------------E---------------Q------X--------C---K-------------------------------------------KH-E-I---------- 349
N.CM.YBF30               ------------------------------------Q-------------C---K-----------------------------------T---------EKH-E-I---------- 356
O.CM.ANT70               ---G-----------------E----------G---Q-Q------------C---P-----------V-------------------------------D---RD--ELE-C----------P 356
O.CM.MVP5180             ------------------------------G---QRQ------------C---P-----------V------V----------------------D----S--EVE----I---------P 345
O.SN.MP1299              ------------------------------G---Q-Q------------C---P-----------V-------------------------------D----D--ELE-C----------P 345
O.SN.MP1300              ------------------------------G---Q-Q------------C---P-----------V-------------------------------D----N--ELE-C----------P 345
CONSENSUS 01             k-d--k-----d------r-q--------i-h-------------------------d-sf----------n----i----------k-------s-m------frik--Em-----m---------- 345
01_AE.CF.90CF11697       E-------------------------V-------------------------------Y-G-------------------------------------------Y-AK--E---------- 349
01_AE.CF.90CF402         ------------------------------------------------------G-----------------V-------------------------------AR--E---------- 347
01_AE.CF.90CF4071        ------------------------------------------------H-SX-------------------------R-------C----------AK--E---------- 349
01_AE.TH.93TH057         ----------------G-------------------------------------S-------------D-------------------------------IK--EM---------- 349
01_AE.TH.93TH065         ------$----G-----------------R-------------------------S-------------------------------------------IK--EM---------- 349
01_AE.TH.93TH253         ------------------------------------------------------S-------------------------------------------IK--EM---------- 346
01_AE.TH.93TH902         ----------------G-------------------------------------S-------------------------------------------IK--EM---------- 341
01_AE.TH.94TH702         ----------------G-S-----------------------------------S-------------------------------------------IK--EM---------- 349
01_AE.TH.95TNIH047       --N--R-----------------------------------------------S-------------------------------------------GIK--EM---------- 349
01_AE.TH.CM240           ------------------------------------------------------S-------------------------------------------IK--EM-----K---------- 349
01_AE.TH.TH022           -----R-----------------------R-----------------------S-------------------------------C-T--------TK--E---------- 354
CONSENSUS 02             kkd--k---------------df--v--g---------k-------------kd----------v-n-------------------a---a--tk---pf-tK--Eiv----------e 349
02_AG.CM.MP807           ------------------------------------R-------------G-----------V-------------------------A----------TK--E----------K 348
02_AG.FR.DJ263           -R-------------------S-----------------------------K-----------T-------------------------A---N---HY-IK--E-M---------- 349
02_AG.FR.DJ264           ---------------------G------R--------------------K-----------V-------------------------A----------IK--E---------- 349
02_AG.GH.G829            ------------------------------------------------------K-------------------------------------------Y-IK--E---------- 349
02_AG.NG.IBNG            ------------------------------------------------------K-----------V-------------------------A----------TK--E---------- 349
02_AG.SE.SE7812          --N--R-------------------I-------------------------------K-------------------------------A----------TK--E---------- 349
02_AG.SN.MP1211          R-------------------------------------------------------V-------------------------------A----------TK--EL--------- 349
02_AG.SN.MP1213          ------------------------------------E-----------------K-------------S-------------------S------I-------AK--E---------- 349
03_AB.RU.KAL153-2        -----------G-----------------------------------------Q-----------T-------------------------------------E---------- 349
03_AB.RU.RU98001         ----------------------------L-----------------------Q-----------T-------------------------------------E---------- 349
04_cpx.CY.94CY032-3      ------------------------------------------------------PE----------T-------------------------C----------FK--E---------- 348
04_cpx.GR.97PVCH         --N--R-----------------------------------------------PA----------T------V------------------Y----------TR--E---------- 348
04_cpx.GR.97PVMY         --N-NR-----------------------------------------------PE-------------------------------------C----------TK--EM--------- 348
05_DF.BE.VI1310          -----R-----------------------------------I-----------KE-------------------------------------C----------EV---------- 349
05_DF.BE.VI961           -----R-----N------K-------------------------------H-------------------F-----------------C----------EM---------- 349
06_cpx.AU.BFP90          ------------------------------------------------I----K----------T-------------------------I-------IK--E---------- 351
06_cpx.ML.95ML127        ------------------------------------------------I---K----------T-------------------------C------D---IK--EL--------- 348
06_cpx.ML.95ML84         ---------------------------------R----------------G-N----------L-------------------------I-------TK--E---------- 350
06_cpx.SN.97SE1078       ------------------------------------------------------K-------------------------------------I-------QK--E---------- 353
10_CD.BFL061             ------------------------------------------------------Y-------------------------------------------QM----------Q 348
10_CD.BFL071             ----------------R--------R-------------------------G-----------------------------------------------EM---------- 349
10_CD.BFL110             ------------------------------------------------------Y-----------------T-------------------------------EM---------- 346
11_cpx.CM.MP818          ------------------------------------------------------N-----------------V-------------------------T---E---------- 350
11_cpx.FR.MP1298         ------------------------------------------------------KE----------T-------------------------C----------E---------- 348
11_cpx.FR.MP1307         --N-------------------------------------------------S-------------------------------------------T----------M----- 364
11_cpx.GR.GR17           ------------------------I-----------------------L---S----------L-------------------------------T---EV---------- 348
CPZ.CD.CPZANT            ---TS--------------------I----------Q----------I---Q----------V---------C----------A------A---DKY-AVE----------M- 349
CPZ.CM.CAM3              ------------------------------------------------C----N----------V-------------------------I---T---Q-H--LI---------- 354
CPZ.CM.CAM5              ---------------------E-------------------------C----N----------V------V-------X----------D-----H--LI---------- 350
CPZ.GA.CPZGAB            ------------------------------------------------C---K----------V-----------S----------EK----T---------- 347
CPZ.US.CPZUS             ---------------------------------------R--------C---KE----------V-------------------------D-----H--VI----------N 351
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins


```
                                    T215Y K219Q
                                        |  |
B.FR.HXB2           IGQHRTKIEELRQHLLRWGLTTPD.KKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSK  478
CONSENSUS A         -g---t------a---s--ft---.k-----------l----------e----e--------------------a----k---k----a----DiVt----------------d----v------  478
A.KE.Q23            -----A------A---S---I----.----------------------------E--------------------A----K--------A----D-VT----------------------------  476
A.SE.SE6594         -----A------S---S--F----.Q---------------------------------E--------------------K---R----A----DIVE----------------D----------  478
A.SE.SE7253         ----------A---S--F----.-------------------------E----E--------------------A----K---R----A----DIVT----------------D----------  473
A.SE.SE7535         ----------E---S--F----.----------------V----------E--------------------A----------A----DIVT----------------D----------  478
A.SE.SE8538         ----------A---S------.--------------------------N----E--------------------A----K--------A----DIVT----------------D----A------  476
A.SE.UGSE8131       -E---------A---S--FF---.Q--------------------------K----E--------------------A----K--------A----DIVT----------------A----------  478
A.UG.92UG037        ----------E---K--F----.--------------------------E----E--------------------A----K----------DIVT----------------D----------  472
A.UG.U455           -----A------A---S--FI---.--------------------------Q--------------------A----K--------A----DIVT----------------D----A------  479
A2.CD.97CDKFE4      -----A------A------F----.--------------------------K--------------------A----K---R---------DIVT--K------E-------N----------  477
A2.CY.94CY017-41    -S---V------A---K--FY---.--------------------------K--------------------A----K--------A----DIVT--K------E-------T----------  470
CONSENSUS B         -----t-i----q---r--f----?----------------p--------v--------------------a----kq-------t------ip--e----------kep--gv----s-  478
B.AU.MBC18          -----I----------K--F----.-------------------------------------------------------------------------------------------R----------  478
B.CN.RL42           -----A----------F----.----------------L--------M--------------------A----KE-----------------------------------S----------  490
B.DE.D31            ----------------F----.--------------------------T--------------------A----------A----------K----------------------------  478
B.GA.OYI            ----------------F----.--------------------------M--------------------A----KN---------------------------------------------  478
B.GB.CAM1           ---------------------.--------------------------M--------------------A----K----------V----------------------------------  478
B.KR.WK             -----I----------F----.-------------------------------------------------A----K----------V----------------------------A-  478
B.NL.3202A21        ----------------F----.-------------------------------------------------A-----H---------------------------------------------  478
B.TW.TWCYS          -----I-V--------F----X--------------------------------------------------A--------------V---K----------Q----A------  479
B.US.JRCSF          ----------K--F----.-------------------------------------------------A----K----------------K----------------------------  482
B.US.JRFL           -----A----------F----.-------------------------------------------------A----K---------------------------------------------  478
B.US.MN             -----A------R------F----.-------------------------------------------------A----K---------------------------------------------  481
B.US.RF             -----I------E---K--F----.-------------------------------------------------A----K----------VQ--K----------------------------  477
B.US.SF2            ----------------F----.----------------M--------------------A----K-------------------------------------E----------  478
CONSENSUS C         ---hrAk---lre---k--f---d.----------------------p-q--e-d----n----------------ik-rq--k----a-a--DiVpl----------------v---y-----  474
C.BR.92BR025        -----A------E---K--F----.--------------------------Q-------------------------------A----DIV-----------------------------  476
C.BW.96BW01B03      ----TA----SKE---K--F----.--------------------------N----E-------------------------A----DIV-----------------------------  474
C.BW.96BW0402       -----A------E---K------Y.--------------------------Q--D------H----------R-KH-------A----DIV-----------------------------  474
C.BW.96BW0502       ---R-AQ-----E---K--F----.--------------------------Q--D-------------------------A----D-V-----------------------------  484
C.BW.96BW1104       -----AR-----K------F----.--------------------------S-K----E----------V----------A----DIV-P----------------------------  472
C.ET.ETH2220        -----AP-----E---K--F----.--------------------------Q-------------------------A----DIVT----------------------F-----  485
C.IN.93IN999        ----A------------F----.--------------------------Q----------------K---R----A-V--DIV-----------------------------  483
C.IN.94IN11246      -----A------E---K--F----.--------------------------Q-------------------------A----DIV----------------I---------  474
C.IN.95IN21068      -----A------K------F----.--------------------------Q-------------------------------DIV-----------------------------  473
D.CD.84ZR085        -----A---K--E----------.----------------------S-T----E--------------------K----------------------------------M---------  479
D.CD.ELI            ---------K--E------F-R--.----------------------S-K----E--------N--ER-------------------------------------------------  477
D.CD.NDK            ----------E------F----.----------------------N----E--------------------A----K----------V----------------------------  477
D.CD.Z2Z6           ----------E------F----.----------------------S-K----E-------------------------------------------------------------  477
D.UG.94UG1141       -----I------G---K--F----.--Y---------------------H----E----------------------------C---A----------A----------------A------  479
F1.BE.VI850         ----------E------F----.------------H-----------Q--N----------------------P-------A----DIV---A-------K-----R----------  478
F1.BR.93BR020-1     ----------E---K-------.----------------------Q--D----------------------K--------A----DIV---T--------------------A------  478
F1.FI.FIN9363       ----------E---K--F----.----------------------Q--D------D----------X----------A----DMV---A--N-------------------------  477
F1.FR.MP411         L----M------E---K--F----.----------------------Q-------------------------IK--------A----DIV-----------------------------  480
F2.CM.MP255         ----------E---K--F----.----------------------Q----S----------RIKH--R----A----D-V---A----------I----------------------  478
F2.CM.MP257         -----A------E------F----.----------------------A-Q--D-S----------R-KH----------D-V---A----------------------------------  478
G.BE.DRCBL          -----A------E------F----.----------------------Q--N-EN----------K-----I--A----DIVSM-A---M----------------------------  475
G.FI.HH8793-1-1     -----A--K---E------F----.----------------------Q----E-------------------------A----DIV---A----------------------------  477
G.NG.92NG083        -----A------E---K-------.----------------------Q----ED----------KH--R----A----DIV---A---M----------H----------------  477
G.SE.SE6165         -----A------E---K--F----.----------------------Q--D-E----------TH-------A----DIVS--A---M---------R----------------  475
```

```
                                    T215Y K219Q
                                    |    |
B.FR.HXB2                IGQHRTKIEELRQHLLRWGLTTPD.KKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSK  478
H.BE.VI991               -----E------A------F-----.Q-----------------------VK-------------------------------K---X-----A-----IV---K-------------------A------  476
H.BE.VI997               -----A------A------F-----.------------------------VK-------------------------------------A----D-V---K-------------R-----------  476
H.CF.90CF056             -----A------A---K--F-----.------------------------TVK-----------------------N---K--------A----DI----K-------------R--I---------  477
J.SE.SE7022              -E---R--K---E---K--FY---.-------------------------Q----ED-----------------------IKE----I--A----DIV---R---------K-------------AR  476
J.SE.SE7887              -E---R--K---E---K--F-----.-------------------------Q----ED-----------------------K------K-A----DIV---R---------K-------------SA-  476
K.CD.EQTB11C             -----A------E------F-----.-------------------------Q--D----------------------F-----K--------V----DIV---A-------------------------  478
K.CM.MP535               ---P--------E---K--F-----.-------------------------Q--D----------------------------K--------V----DIV---A-------------------------  478
N.CM.YBF106              -A---ETV----G---K--F-----.-------------------------K----EV----------------------K-----I---------VTF-Q---------------L-------G-  485
N.CM.YBF30               LA---EAV-D--D---K--F-----.-------------------------K-----V----------------------R-K-----I---------VNF-----------------L-------G-  485
O.CM.ANT70               LTE--KR--L--E--YQ--F-----.-------------------------S-Q--N--V---------I----------Q--R--E----I----S----V--SR------E----R--Q-------Q-D-  474
O.CM.MVP5180             LAE--KRV-L--E--YQ--F-----.-------------------------Q--D-EV---------------------Q--R-KE----I----S----V--SK------E----K----------Q-D-  474
O.SN.MP1299              LTE--KRV-L--E--YQ--F-----.-------------------------Q--N-EE---------------------Q--R-KE----I--A-S---IV--SK------E----K----------Q-D-  474
O.SN.MP1300              LTE--KRV-L--E--YQ--F-----.-------------------------Q--N-EE---------------------Q--R-KE----I----S----V--SK------E----K----------Q-D-  474
CONSENSUS 01             i-q-rt-i-e--a---s--F---d.k-----p----------R------e--e-d----n-------------a----k----l---a----diVp---------a------kt----------  478
01_AE.CF.90CF11697       -----------D---K--F-----.------------------R-------E-------------------------A----K--------A------V-----------------N----------  476
01_AE.CF.90CF402         -------V-D--A---S--F-----.------------------R-------------------------------A-------------A----DIVT----------------T----------  478
01_AE.CF.90CF4071        ----G-------A---S--F-----.------S----------R------Q--D-E----D---------------A----K--------A----DIV-----------------N----------  478
01_AE.TH.93TH057         -----------A---S--F-----.------------------R------E-------------------------A----K----P--------DIV-----------------T----------  478
01_AE.TH.93TH065         -----I------A---S--F-----.------------------R------E-------------------------K--------A----DIV-----------------RT----------  475
01_AE.TH.93TH253         -----------A---S--F-----.------------------R------E-------------------------A----K----------DIV----------E------RI----------  470
01_AE.TH.93TH902         T-P--A------A---S--F-----.------------------R------E-------------------------A----K--------A----D-V-----------------RT----------  478
01_AE.TH.94TH702         T----A------A---S--F-----.------------------R------E-------------------------A----K--------A----D-V-----------------RT----------  478
01_AE.TH.95TNIH047       -----------A---S--F-----.Q-----------------R------E-------------------------A----K--------A----DIV-----------------T----------  478
01_AE.TH.CM240           -----------A---S--F-----.------------------R------E-------------------------A----K--------A----DIV-----------------T----------  483
01_AE.TH.TH022           -----------A---S--F---G.------------------R------E-------------------------K--------A----DTV-----------------T----------  478
CONSENSUS 02             -g---A------e---?--F----.k----------------------iq--e----------l--asq-ya---vKq--k----a-a--DIVt-------------------------t-  476
02_AG.CM.MP807           -----A------E------F-----.-------------------------Q-------------------------A----K--------A----DIVT----------------------A-  478
02_AG.FR.DJ263           -E---A------E---K--F-----.-------------------------Q-------------------------A----K--------A----DIV-----------------------A-  478
02_AG.FR.DJ264           -----A------E---K--F-----.-------------------------Q----------------------FA----K---R----A----DIV-----------------------A-  478
02_AG.GH.G829            -----A------E------F-----.-------------------------M-------------------------A----KE-------A----DIVT----------------------T-  478
02_AG.NG.IBNG            -----A------G---K--F-----.-------------------------VE------------------------A---IK---R----A----DIVA----------------------T-  478
02_AG.SE.SE7812          -----A------E------F-----.E-------------------------------------------V--QVR--A----K--------A-T--DIVT----------------------T-  478
02_AG.SN.MP1211          -----A------E---K--F-----.-------------------------Q-------------------------K-------------DIV-----------------------T-  478
02_AG.SN.MP1213          -----A------E------F-----.-------------------------Q--D----------------------A----K--------A----DIVT----------------------T-  478
03_AB.RU.KAL153-2        ------E-----E------F-----.-------------------------------------------------A-------------A----------A-----------------------  478
03_AB.RU.RU98001         -----I------D------F-----.-------------------------------------------------A-------------A-----I----A-------------V----------  478
04_cpx.CY.94CY032-3      -----A------E------F-----.-------------------------QPA-------------------------K--------A----DIV---T-----------------A------  477
04_cpx.GR.97PVCH         ------------E------FY---.-------------------------Q-------------------------K------------DIV---T-----------------A------  477
04_cpx.GR.97PVMY         ----KA------E------FF---.Q------------------------Q-V--E----------------------A----K--------A----DIV---T-----------------A------  477
05_DF.BE.VI1310          -----A--K---E---A--F-----.-------------------------Q--D----------------------K-------------------------------------------  478
05_DF.BE.VI961           ------------E---K--F-----.-------------------------Q--N----------------------K-------------V-----------------------A-  478
06_cpx.AU.BFP90          -----A------E---K--------.-------------------------Q--D-E----------------------------A----DIV---A-------------------------  480
06_cpx.ML.95ML127        -----A------E---K--F-----.-------------------------Q--N-E--SI---------------S-------------A----DIV---A-----------------A------  477
06_cpx.ML.95ML84         -----A------E---K--F-----.-------------------------Q--D----------------------KH--------A----DIV---A-----------------------L-  479
06_cpx.SN.97SE1078       -----E-V----A---K--F-----.-------------------------Q--D----------------------K--------A----DIV---A-------------------------  482
10_CD.BFL061             --H-TI------G---K--F-A--.----------------------L---Q--K-N----------------------------A-V--DIV-------------------------------  477
10_CD.BFL071             -----A---D--G---K--F-----.-----K--------------------Q----E-------------------------------DIV-------------------------------  478
10_CD.BFL110             -----I------E---K--F-----.------------------------S-Q----------------------R----------A----DIV-------------------------------  475
11_cpx.CM.MP818          -----E-V----K------------.-------------------------Q--D-EC---------------------K------------DIV---A-------------------------  479
11_cpx.FR.MP1298         -----E------K---K--F-----.-----------------------T-Q--D-E--------------------S---IK-------------DIV---A-----------------A------  477
11_cpx.FR.MP1307         -----E-V----E--WK--FY---.-------------------------Q--D-EC---------------------K------------DIV---A-----------------------A-  493
11_cpx.GR.GR17           -----E-V----K---K--F-----.-------------------------Q--D-E----------------------A----K--------A----DIV---A-------------------  477
CPZ.CD.CPZANT            -TA--EM--K-----QV---E---.---------Q--------------K-K---P-D----------------------TK---R-I--V-S--DRVQM-R------E--KQ--QQKIE-Y--Q-GL  478
CPZ.CM.CAM3              LRE--E-V-L------T--F----.-------------------------Q--Q-ET---------------------K-----IK------DIVTM-Q---M--E--X---XD---------E-  483
CPZ.CM.CAM5              --K--E-V-L------T--F----.-------------------------Q--Q-EI---------------------S----K-----IK--------VAM-Q---M--E-------D---------E-  479
CPZ.GA.CPZGAB            -D---K-V--------K--F----.-------------------------Q----EV---------I-------------IK-----I----K--D-V---P------------VST---------D-  476
CPZ.US.CPZUS             LEK--E-V-L---Y--T--F----.----E----------------------Q--Q-EI-----------------------IK-----IK-A------VNF-H---M--E----------------E-  480
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins


```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins


```
B.FR.HXB2          MASDFNLPPVV.AKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLA.GRWPVKTIHTDNGSNFTGATVRAACWW.AGIKQEFGIPYNPQSQ  863
CONSENSUS A        ---------?v.--------d--------m---------i-------------l---------i------------t---l----.------vv---------S-AvK-a---.-n?q--f---------  857
A.KE.Q23           ---------I-.-------------------------------------------V-----------------------------.------IV---------S-A-K-----.-N-Q-----------  863
A.SE.SE6594        ----------I.-----------------I-----------------------------------------------------.------VV---------S-A-K-----.-N-Q-----------  858
A.SE.SE7253        ----------I.-------------------------------------------------------------------------.------IV---------S-AFK-----.-S-Q-----------  863
A.SE.SE7535        ---------I-.-------------------------------------------------------------I----.------V----------S-A-K-----.-NV------------  861
A.SE.SE8538        ----------I.--------N-----------------M-----------------------------------------.------VV---------S-AFK-----.-NVQ--Y---------  863
A.UG.92UG037       ---------I-.-------------------------------------------------------A--------.------VV---------S-AFK-----.-NVQ-----------  857
A.UG.U455          ----------.--------N-------------------------------------V---------------------.------VV---------S-A-K-----.-NV------------  863
A2.CD.97CDKFE4     --Q------I-.-------------------------------------------------------------I----.------V----------S-A-K-V---.-N-Q-----------  862
A2.CY.94CY017-41   --H--------.-------------------------------------------------------------I----.------V------P---S---K-----.--VQ-----------  855
CONSENSUS B        --s------v-.------s-------------------------------T----D----I----.------V------P--IS---K-----.---Q-----------  863
B.AU.MBC18         ---------I-.-------------------------i-------------------------l----.-----kt------s--tStt-K-----.agi--e--i-------  863
B.CN.RL42          ----------.-------------------------------I-------------------------I----.-----T----------IST--K-----.--------T-------  875
B.DE.D31           ----------.-------------------------------------------------------------.-------------R---SNS-K-----.---------------  863
B.GA.OYI           ----------.-------------------------------------------------------------I----.----------------ST--K-----.--V-----------  863
B.GB.CAM1          ----------.-------------------------I-----------------------------------I----.----------------ST--K-----.---------------  863
B.KR.WK            --G--------.------C-----------------------------------------------------------.--------------G--IST--K-----.---------------  863
B.NL.3202A21       ----------.-------------------------------------------------------------------.----------------SN--K----$.-R-------------  862
B.TW.TWCYS         ----------.-------------------------------I-----------------------------I----.-----T----------S---K-----.---------------  863
B.US.JRCSF         ---------I-.-------------------------------I-------------------------------.----------------S-A-K-----.-----K----------  864
B.US.JRFL          ----------.-------------------------------I-------------------------------.-----T----------ST--K-----.---------------  867
B.US.MN            ---------I-.-------------------------------------------------------------.----------------ST--K-----.---------------  863
B.US.RF            ----------.-------------------------------------------------------------.-------------P---ST--K------.T--------------  865
B.US.SF2           ----------.-------------------------------I-------------------------I----.------V----------ST--K-----.---------------  862
CONSENSUS C        --se--l--i-.a-----s-dk--l----i----d-s-------------i----------i----------fi----.-----kv------s--tSaa-K-----.---Q-----------  863
C.BR.92BR025       ---E-----I-.----------------T-------------------I-------------------------I----.------V----------ISN--K-----.---Q-----------  859
C.BW.96BW01B03     ---E-----I-.----------------------G-------------------------M----------YI----.------V------T---S-A-K-----.---Q-----------  860
C.BW.96BW0402      ---E-----I-.----------------I-------------------I-------------------------YI----.------V----------STA-K-----.---Q-----------  859
C.BW.96BW0502      --DE-----I-.----------------I-------------------------------M----------I----.-----RV------T---S-A-K-----.---Q-----------  859
C.BW.96BW1104      ---E-----I-.----------------I-------------------------------------------YI----.------I----------S-A-K-----.---Q-----------  869
C.ET.ETH2220       --NE--I----.P-----C----------I----N---------------I-------------------------.-----RV----------SNA-K-----.---Q-----------  857
C.IN.93IN999       ---------I-.--------GQ--Q-----------------------I-------------------------YI----.------V----------S-A-K-----.---Q-----------  870
C.IN.94IN11246     ----------.---------Q-------------------------I----------M----------------I----.------V----------S-A-K-----.---Q-----------  868
C.IN.95IN21068     ----------.---------Q-------T-----------------I----------M----------------I----.------V----------S-A-K-----.---Q-----------  859
D.CD.84ZR085       ----------.-------------------------------------------------------A--------.------VV---------S---K-----.---Q-----------  858
D.CD.ELI           ----------.-------------------------------------------------------------.------VV---------S-A-K-----.---------------  864
D.CD.NDK           ----------.-------------------------------------------------------------.------VV---------S-A-K-----.---------------  862
D.CD.Z2Z6          ----------.-------------------------------------------------------------I----.------IV---------S-A-K-----.---------------  862
D.UG.94UG1141      ---E-------.------------V----L-------------------G-------------------------.------VV---------S-A-K-----.---------------  862
F1.BE.VI850        ---------I-.-------------------------------------------------E----------I----.------I----------S-A-K-S---.---Q-----------  864
F1.BR.93BR020-1    ------I-A--.-------------------------------I-----------L-----------------.--------------T---S---K-----.---Q-----------  863
F1.FI.FIN9363      ----------.-------------------------------I-----------------------D----I----.------M----------S-A-K-----.---Q-----------  862
F1.FR.MP411        ----------.-------------------------------I-------------------------I----.------I----------SSA-K-----.---Q-----------  865
F2.CM.MP255        ----------.------N------------------------I-------------------------I----.------I----------STV-K-----.---Q----V-------  863
F2.CM.MP257        ----------.-------------------------------I----------------------A---I----.------I----------S-V-K-----.---Q-----------  863
G.BE.DRCBL         ---------I-.-------------------------------I-------------------------I----.------I----------S-A-K-----.-S-T-----------  860
G.FI.HH8793-1-1    --------HI-.--------H----------------------I-------------------------I----.------V----------S-A-K-----.-D-T-----------  862
G.NG.92NG083       ----------.-------------------------------I-I-----------------------I----.------V------P--IS-A-K-----.-N-T-----------  862
G.SE.SE6165        ---------I-.-------------------------------I-I-------------------------I----.-----TV----------S-A-K-----.-N-T-----------  860
```

```
B.FR.HXB2             MASDFNLPPVV.AKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLA.GRWPVKTIHTDNGSNFTGATVRAACWW.AGIKQEFGIPYNPQSQ  863
H.BE.VI991            ---------I-.-------------------------------------------I-------------------------------I----.------M----------S-A-K-----.-D-H-----------  861
H.BE.VI997            ---E-----I-.-------------------------------------------XP------------------------------I----.------M------T---STA-K-----.-D-Q-D---------  860
H.CF.90CF056          V--------I-.-----------------------------------------Q-----------------------K----------.S-----V----------S-A-K-----.-D-Q-----------  862
J.SE.SE7022           ----------.-------------------------------------------------------------------A-F-I----.-G----A----------SGA-K-----.-D-------------  861
J.SE.SE7887           ----------.-------------------------------------------------------------------A-F-I----.------V----------SGA-K-----.-D-------------  861
K.CD.EQTB11C          ----------.----------------I---------------------I-------------------------------I----.-----RV----------S-V-K-----.-D-------------  864
K.CM.MP535            ---------I-.----------------I---------------------I-------------------------------I----.------V------T---STV-K-----.--V------------  863
N.CM.YBF106           ---------I-.--------------------IN----V----------I----------L----------------I----.------V------P--IS---K-----.----------------  870
N.CM.YBF30            ---------I-.----------------------N----V----------I----------L----------------I----.------V----------S---K-----.-N-------------  870
O.CM.ANT70            L--E-G-----.----I---P--HI----I---------EV--I----M---I-I--------F-------------------.A-----V------P---ST-MK-----.-N-QH----------  859
O.CM.MVP5180          L----G---I-.----I---P--HI----T-----Y--E---M--------I-I-------DF-------------------.A-----V------P---S-AMK-----.T--QH----------  859
O.SN.MP1299           L----G-----.----I-N-PQ-HI----I--------EV--M--------I-I--------F-------------------.A-----V------P---S-AMK-----.-N--H----------  859
O.SN.MP1300           L----G-----.----I-N-P--HI----I-----Y--EV--I--------I-I--------F-------------------.A-----I------P---S--MK-----.T---H----------  859
CONSENSUS 01          ---------i-.-----an---------am-------------------------------v--g-------------------.-----kVi----------S-avK-----.-NVr---------s-  863
01_AE.CF.90CF11697    ---------I-.------N-----------------------------------------------------------------.--$---V----------S-AMK-----.-NVQ-----------  860
01_AE.CF.90CF402      ---------I-.-------------------------------------------------------------------.-----RV----------S-A-K-----.-NVQ-----------  863
01_AE.CF.90CF4071     ---------I-.------N-----------------------------------------------------------------.------V----------S---K-----.-NVQ-----------  863
01_AE.TH.93TH057      ---------I-.------N-----------------------------------------------------------------.------V----------S-A-K-----.-NVR----------G-  863
01_AE.TH.93TH065      ---------I-.------N-----------------------------------------------------------------.------V----------S-A-K-----.-NVR-----------  860
01_AE.TH.93TH253      ---------I-.------N---------P-------------------------L-----------------------------.------V----------S-A-K-----.-NVR-----------  855
01_AE.TH.93TH902      ---------I-.------N---------PX------------------------X-----------------------------.------V----------S-A-K-----.-NVR-----------  863
01_AE.TH.94TH702      ---------I-.----------N-----------------------------------------------------------.------V----------S-A-K-----.-NVR-----------  863
01_AE.TH.95TNIH047    ----------.-----TN-------------------------------------------------------------------.------V----------S-A-K-----.-NVR-----------  863
01_AE.TH.CM240        ---------I-.------N-----------------------------------------------------------------.------V----------S-A-K-----.-NVQ-----------  868
01_AE.TH.TH022        ---------I-.---d-----iv.------s--k--l-----------s-----l--------i-l-------------------I----?----v-vi---------S-A-k-----?anvt--f--p------  860
CONSENSUS 02          ---d-----iv.------s--k--l-----------s-----l--------i-l-------------------I----?----v-vi---------S-A-k-----?anvt--f--p------  860
02_AG.CM.MP807        ---------I-.-------------------------------------------I-M-----------------I----.------V----------S-A-K-----.-NVT-----------  863
02_AG.FR.DJ263        ---------II.------C--Q-------------G-------------------I-------------------I----.------V----------S-A-K-----.-NVT-----------  863
02_AG.FR.DJ264        ---------I-.------C-----------------------------------I-------------------I----.------V----------S-A-K-----.-NVT-----------  863
02_AG.GH.G829         ---------I-.-------------------------------------------I-------------------I----.------I----------S-A-------.-NVT-----------  863
02_AG.NG.IBNG         ---------I-.------------M----------G-------------------I-------------------I----.------V----------S-A-K-----.-NVT-----------  863
02_AG.SE.SE7812       ----------.-------------------------------------------I-------------------I----.------V----------S-A-K-----.-NVT-----------  863
02_AG.SN.MP1211       ---E-----I-.-----------------------------I--------I-I-------------------------I----X----X-VV---------S-A-K-----XPKLH--L--X------  865
02_AG.SN.MP1213       ---------I-.-------------------------------------------------------------------I----.------V----------S-A-K-----.-NV------------  863
03_AB.RU.KAL153-2     ----------.-------------------------------------------I-------------------V----.------I----------STA-K-----.----------------  863
03_AB.RU.RU98001      ----------.-------------------------------------------I-------------------V----.------I----------ISTA-K-----.----------------  863
04_cpx.CY.94CY032-3   --------S--.--------N-----------------------------M-------------------------I----.------M--A---P---S-A-K-----.-D-N-----------  862
04_cpx.GR.97PVCH      ----------.--------N-----------X-----------------I---P-------------------------I----.------I------P---S-A-K-----.-DVQ-----------  861
04_cpx.GR.97PVMY      ----------.--------N-------------------------------I-------------------------I----.------I----------S-A-K-----.-N-Q----V-------  862
05_DF.BE.VI1310       ----------.-------------------------------------------------------------D----I----.------M------P---SGA-K-----.---Q-----------  863
05_DF.BE.VI961        ----------X-------------------------------------------------------------------I----.------MV---------S-A-K-----.----------------  864
06_cpx.AU.BFP90       --N------I-.-----------------I---------------------I-------------------------I----.------V----------S-A-K-----.-N-T-----------  865
06_cpx.ML.95ML127     ---------I-.-------------------------------------------I-------------------I----.------V----------S-A-K-----.-N-T-----------  862
06_cpx.ML.95ML84      ---------IL.------C-----------------------------------I-------------------I----.------V----------STA-K-----.-NVT-----------  864
06_cpx.SN.97SE1078    --N------I-.----------------P---------------------I-------------------------I----.------V----------S-A-K-----.-NVT-----------  867
10_CD.BFL061          ----------.----------------L-------------------------------------------------.------VV---------S-A-K-----.----------------  862
10_CD.BFL071          ----------.----------------L-------------------------------------------------.------VV---------S-A-K-----.----------------  863
10_CD.BFL110          ----------.----------------L-------------------------------------------------.------VV---------S-A-K-----.----------------  860
11_cpx.CM.MP818       ---------I-.-------------------------------------------I-M----------------T---------I----.------V----------SS--K-----.-N-Q-----------  864
11_cpx.FR.MP1298      ----------I.-------------------------------------------------------------------I-R--.------M----------S-A-K-----.-N-Q-----------  862
11_cpx.FR.MP1307      ---------I-.-------------------------------------------I-------------------------I----.------V----------S-A-K-----.-N-Q-----------  878
11_cpx.GR.GR17        ----------I.------N-------------------R------------I-------------------------I----.------V----------SNA-K-----.---Q-----------  862
CPZ.CD.CPZANT         L-DEY----I-.----I-Q----HV----R----------V----------I-----S--F-----MAD---KS---------.S----------A---S-A-K-----.-N-Q-----------  863
CPZ.CM.CAM3           ------I-HI-.---L--R-----------------------------------------L----------------I----.------V------P---SN--K-----.---Q-----------  869
CPZ.CM.CAM5           ----------.-------------------------------------I----------L----------------I----.------V------P---SN--K-----.---Q-----------  864
CPZ.GA.CPZGAB         ---------I-.------H-----V----------V----------I-------------------------------.-------------P---S-A-K-----.-D-------------  861
CPZ.US.CPZUS          ----------.----------P----I-------------------I-I---------L----------------I----.------V----------SS--------.---Q-----------  865
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins


```
                                                
```

```
                                                
```

```
B.FR.HXB2              GDDCVASRQDED$          1003
CONSENSUS A            ----v-g------           997
A.KE.Q23               ------G------          1003
A.SE.SE6594            ------G------           998
A.SE.SE7253            ------G------          1003
A.SE.SE7535            ------G------          1001
A.SE.SE8538            ------G------          1003
A.SE.UGSE8131          ------------           997
A.UG.92UG037           ------G------          1003
A.UG.U455              ----M-G------          1002
A2.CD.97CDKFE4         ------C------           995
A2.CY.94CY017-41       ------G------          1003
CONSENSUS B            ------s------          1003
B.AU.MBC18             ------------          1015
B.CN.RL42              ------------          1003
B.DE.D31               ------G------          1003
B.GA.OYI               ------------          1003
B.GB.CAM1              ------------          1003
B.KR.WK                ------------          1002
B.NL.3202A21           ------------          1003
B.TW.TWCYS             ------G------          1005
B.US.JRCSF             ------------          1007
B.US.JRFL              ------------          1003
B.US.MN                ------------          1005
B.US.RF                ------------          1002
B.US.SF2               ------------          1003
CONSENSUS C            -A--v-gr---d-          999
C.BR.92BR025           -A--M--------         1000
C.BW.96BW01B03         -A----G------          999
C.BW.96BW0402          -A----G------          999
C.BW.96BW0502          -A----GG---N-         1009
C.BW.96BW1104          -A----G------          997
C.ET.ETH2220           -A----G------         1010
C.IN.93IN999           -A----G------         1008
C.IN.94IN11246         -A----G------          999
C.IN.95IN21068         -A----G------          998
D.CD.84ZR085           --------------        1004
D.CD.ELI               ------------          1002
D.CD.NDK               ------------          1002
D.CD.Z2Z6              ------------          1002
D.UG.94UG1141          ------------          1004
F1.BE.VI850            V-----G------         1004
F1.BR.93BR020-1        ------G------         1003
F1.FI.FIN9363          ------G------         1002
F1.FR.MP411            ------G------         1005
F2.CM.MP255            ------G------         1003
F2.CM.MP257            ------G------         1003
G.BE.DRCBL             ------G------         1000
G.FI.HH8793-1-1        ------G------         1002
G.NG.92NG083           -G----G------         1002
G.SE.SE6165            ------G------         1000
```

```
B.FR.HXB2              GDDCVASRQDED$          1003
H.BE.VI991             ------G------          1001
H.BE.VI997             ------G------          1000
H.CF.90CF056           ----------             1002
J.SE.SE7022            ------G------          1001
J.SE.SE7887            ------G------          1001
K.CD.EQTB11C           ------G------          1003
K.CM.MP535             ------G------          1003
N.CM.YBF106            --G----GX--NQ          1011
N.CM.YBF30             --G----G---NQ          1011
O.CM.ANT70             -T-SM--G-T-SE          1000
O.CM.MVP5180           -T-SM-N--T-SE          1000
O.SN.MP1299            -T-SM--G-T-SE          1000
O.SN.MP1300            -T-SM--G-T-SE          1000
CONSENSUS 01           ------g--d-d-          1003
01_AE.CF.90CF11697     ------G----N-          1000
01_AE.CF.90CF402       ------G--N---          1003
01_AE.CF.90CF4071      ------G------          1003
01_AE.TH.93TH057       ------G------          1003
01_AE.TH.93TH065       ----------             1000
01_AE.TH.93TH253       ------G------           995
01_AE.TH.93TH902       ------G------          1003
01_AE.TH.94TH702       ------G------          1003
01_AE.TH.95TNIH047     ------G------          1003
01_AE.TH.CM240         ------G------          1008
01_AE.TH.TH022         ------G------          1003
CONSENSUS 02           ----v-gr-d---          1000
02_AG.CM.MP807         ----------             1003
02_AG.FR.DJ263         ------G------          1003
02_AG.FR.DJ264         ------G------          1003
02_AG.GH.G829          ------G------          1003
02_AG.NG.IBNG          ------G------          1003
02_AG.SE.SE7812        ------G------          1003
02_AG.SN.MP1211        ------GG-N---          1005
02_AG.SN.MP1213        ----M-G------          1003
03_AB.RU.KAL153-2      ----------             1003
03_AB.RU.RU98001       ----------             1003
04_cpx.CY.94CY032-3    -N----G------          1002
04_cpx.GR.97PVCH       ------G------          1001
04_cpx.GR.97PVMY       ------G------          1002
05_DF.BE.VI1310        ------G------          1003
05_DF.BE.VI961         ------G------          1005
06_cpx.AU.BFP90        ------G------          1005
06_cpx.ML.95ML127      ------G------          1002
06_cpx.ML.95ML84       ------G------          1004
06_cpx.SN.97SE1078     ------G------          1007
10_CD.BFL061           -A----------Q          1003
10_CD.BFL071           -A----------Q          1004
10_CD.BFL110           -A----------Q          1001
11_cpx.CM.MP818        ------G------          1004
11_cpx.FR.MP1298       ------G-----Q          1003
11_cpx.FR.MP1307       ------G----N-          1018
11_cpx.GR.GR17         ------GS-----          1002
CPZ.CD.CPZANT          DR-DL-G--N---          1003
CPZ.CM.CAM3            ---SM-GG---SQ          1010
CPZ.CM.CAM5            --GSM-G--N--Q          1005
CPZ.GA.CPZGAB          ---------N---          1001
CPZ.US.CPZUS           -A-SM-G----S-          1005
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

<- pol cds end

Hydrophilic region

```
B.FR.HXB2          MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHYESPHPRISSEVHIPLGD..ARLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPELADQLIHLYYFDCFSDSAIRKALLGHI  128
CONSENSUS A        m-----vmi----------N-------y?-kk-kg-f----y-sRh-kv-------l-d..-rlvvrt-w--htg-kd-h--hg----w-l?r----id-Dl---l--lh-fd---d------il-qv  128
A.BY.97BL006       X------X----------N----------K---E$V-------RQ--V----------..----VK--X---AX-K--Q--HX----X-QEX----I--D----------E------IV---  126
A.DE.AF200476      ---------------------N----------K--K-------F-CR--KA----------..----VR----------------------Q---------D---------------------I--E-  127
A.KE.AF233689      ---------------------N--------I-K--KN---------R--KV----------E..----VR-----Q---K-----H-------Q------I--D--------H------E------I--QV  128
A.KE.Q23           ------A----------N-------H--K--KR---------R--KV----------..-T--VRA--------K-----H-------L------IT-D-------MH-------------IV-QV  128
A.SE.SE6594        ---------------------N----------K--KN-V----F--R--KV----------..-K--VR---------KE----H-------LN-----I--DQ-------H----------QV  128
A.SE.SE7253        ---------------------N----------R--KD---------RN--T----------..----VR---------K--Q--H-------LR-----I--D----Q---H----------I---V  128
A.SE.SE7535        ---------------------N-------C--K---N---------R--KV----------..----VR-----Q---K-----H-------L------I--D----------L----------QV  128
A.SE.SE8538        ---------------------N--------I-K--KN------F--R--KV---------E..-K--VR-----Q---K--Q--H-------LR-----I--D-------MH-XN---------I--QV  128
A.SE.SE8891        --------V------------N--------I-K--K--L----F--R--KV---------E..----VR-----Q---K-----H-------LR-----I--D--------H--K----------I--E-  128
A.SE.UGSE8131      ---------------------N--------I-K--A----------R--KV----------..---IVR------P-----Q--H-------LR-----I--D--------H--N----------I--RV  128
A.UG.92UG037       ---------------------N--------I-RR-K----------R--KV-------I--..--I-VR-----Q---K-----H-------L------I--D--------H--N----------I--QV  128
A.UG.U13-2         ---------------------N--------I-K--K----------R---V----------..--I---------K--K--Q--H-------LR-----I--D--------H--N----------QV  128
A2.CD.97CDKFE4     ----------H----------N------I---R--KD-S-------R--KV---------E..----VR-----QP--QA----H-------Q------I--D---H----------E------I--E-  128
A2.CD.97CDKS10     ---------------------N--------I-K--KK---------R---V---------E..-T-IVR------P--KE-Q--H-------Q------I--D---H----------E------I--E-  128
A2.CD.97CDKTB48    ---------------------N----------K---E---------R---V--------KE..---IVR------P--K-----H-------QG-----I--D---H----------E----R-I--E-  128
A2.CY.94CY017-41   ---------------------N--------I-R--K--V-K-----RN------------E..--IIVR------I--K-----H-------QN--H--I--D---H----------E------II-E-  128
CONSENSUS B        ---rw--m----v-----rt-k---k--m-isk--kgwf-r---esth-ri--e-h--l-d..ar--it------t---d-hl-q-vsi-w-kkr-s--v--dl--q---ly-----se---rnail-hi  128
B.AU.MBC18         ---------------------N----------K--K--I------NT--K----------..-----------------------------ERT-R-------------IH----------A-I---R  128
B.CN.RL42          -------------------K----------I-R--K----K---D-T--K----------..-------------------------------------G---------E----N-I--RV  128
B.DE.D31           ------------------------------EK---K-----TN----------------..-----------------Q--H----------------G---------E----N-I--R-  128
B.ES.89SP061       ---------------------------------R--K----------T----------N..-K-----------------------L--R-----L--N--------H------E----N-I----  128
B.FR.vi02011A1H    -------------------A----------KT----K-G-----D----K----------..-K-----------------------R--------N---H----------E----N-I--YR  128
B.GA.OYI           ------------------------------K--K----------T----------..-T--V----------E------A--------G-------T-------E----N-I--N-  128
B.GB.CAM1          ----------------------------I----KK-S-------T----------E..----V-----------------------T-G-N-----D----------E------IV-RL  128
B.JP.D70887        ----------------------------I-R--K----------T----------..-K---I-----------------------------G----------------N-I----  128
B.JP.PT7-6         ----------------------------I----K--C----------V--A--------..S----------N----------R--------N----------E----N-I--RR  128
B.KR.WK            -------------------K----------I-K--KE-V-------T----------..-K----------E----------N-----D---K----H----------H-I---R  128
B.NL.3202A21       -------------------A----------K-K--ER---------T----------V-E..-----------------------------G---------E----N-I---V  128
B.TW.TWCYS         -------------------A----------I-K--K--L-K-----T----------..-T-----------------------R--------DQ----------E------IV-CR  128
B.US.1-2           ------------------------R------K--KE-S-------T--K-----Q-----..----------E----------E-----I--D--------H------T-----N-I--RL  128
B.US.AF019528      ---S--------------K--H------I-I-E--K--I-K-----NN-K----------E..-K---A----------F----------RK-------D-------Q-------E----N-I--Q-  128
B.US.BC            ---------------------I--------I-R--K----------T--K----------..----------------H---V----------D---------E----N-I----  128
B.US.JRCSF         ---------------------N--------I----K--I-K-----TN--V----Q-----..----------------M---TR--------D---------E----N-I----  128
B.US.JRFL          ------------------------------T----K--I-------T----------..----------------V-------N-----D---------E----N-I----  128
B.US.MN            ----R-------A----------------I-K--K-R--------T----------..-----------------------------D---H----H----------I---R  128
B.US.RF            ------------------------------I-R--K----------T----------P--..E----------R--------D---------E-----PS----  128
B.US.SF2           ------------------------------I-K--K----------T---V----------..-K----------E-------A------K-------G--------H------E---KN-I--YR  128
B.US.WC001         -------K-----------------------------KK---------T----------..-N----------S-----------------G---------E----N-I---R  128
CONSENSUS C        -------L--------k----n-------yv-rr-n--f----ye-Rh-kv----------e..-r-v-k-----q----d----H------rlRr-s-qvd-G---q---MH-----ad------i--hi  128
C.BR.92BR025       -------L--------K----N----------RR-S--Y-------R--K----------E..------I-----Q---------H-------LR----R---G-------MH-----A-------I---R  128
C.BW.96BW01B21     -------L----------N----------RR-S-------F--R--KV----------E..------K-----Q----E----H-------LR--------G-------MH-----A-------I----  128
C.BW.96BW0402      -------L--------K----N----------R--N----------R---V----------..------K-----Q---------H-------LR--------G-------MH-----A-------I--Q-  128
C.BW.96BW0504      ----$--L--------K----N-------HI-KR-K----------R--KV----------E..------K-----Q---------H-------LRK-----E-G-------MH-----A-------I--E-  127
C.BW.96BW1104      -------L-------------T----------RR-N--S----F--R--KV----------E..------K----------------H------TVRG---$---G-------MH-----T----------QV  127
C.ET.ETH2220       -------L--------K----N-------HI-RR-N--V-----D-R--KV----------E..----I-K-----Q---------H-------LRS-N-----G---H---MH-----AE------I--YR  128
C.IN.94IN11246     -------L--------K----N----------RR-N--------D-RN-KV----------E..-I---K-----Q---------H-------LR--N--IE-G-------MH-----A-------I--R-  128
C.IN.95IN21068     -------L--------K----N----------RR-N----------R--KV----------E..-----------Q---------H-------LRK-----E-G-------MH-----A-------I----  128
CONSENSUS D        ----------------------rt-----v-h-myv-K--?g-f-----dsphpk----------e..-r--vk-----h---re----------rr-s--v--G---q---m-y----aes-----i--hi  127
D.CD.84ZR085       -------------------N-------Y--HI-K--K--------D----K----------..----V----------E----------R--------G-------M------A-------I----  128
D.CD.ELI           -------------------K----------K--NR----------K----------E..-----K----------E----------R--------G-------M-------E------I--D-  128
D.CD.NDK           -------------------N-------Y-----K--NR-------D-H--K----------E..----V----------KE----------R--------G-------M------AE------I----  128
D.CD.Z2Z6          ------------------------------K--SR-------D----K----------E..-M--VK----------------------R--------G-------M-------EA-----I----  128
D.UG.94UG1141      ----------------------------I-K--K--L-----DC-N-K----------E..----VK----------E----------G--N--I--G-------I-H----AE------I--QV  128
D.UG.U18-0         ------------------------------K--Q--W-----D----K----------E..----VK----------E----------E-------G-------M------AE------I----  128
D.UG.U25-6         -------------------I--------IH--K--T--L-----DCL--K----------E..----VK----------E----------R--------G-------I------AE------I----  128
D.UG.U36-0         ----------------------------Y---I-K--Q--V-----DC-NSK----------E..----VK----------E----------R--------G-------M------AE----------Q-  128
D.ZR.AF233690      -------------------N-----I-Y-----K--S----------------------E..-K--VI-----N----------G--------G---R---M-------E------I--Q-  128
F1.BE.VI850        ------L----------N-------Y-----K--K--S----FQ-R---V---------EE..VK----------P---E----------QGK-R--I--G-------I-------E------I---R  128
F1.BR.93BR020-1    -------------------N-------Y--HI-K--K--------F--R--K---------ET..-E---------LP---E----------QG--R--I--G-------I-------E------I---K  128
F1.DE.AF200475     ---I-------------NQ------Y--H--K--K-------F--R--KV---------KE..-K----------------------------QG--R-----G-------I-------------R  128
F1.FI.FIN9363      -------T----------N-------Y--H--K--KR------F--R--K---------EE..-E---------N----------------QG--R--I--G-------I-------E------I---R  128
F1.FR.MP411        -------------------S-------Y--H--K--KN------FQ-R--K---------EE..-K--V----------------------QG--R--I--D-------I-------A------I---R  128
F2.CM.MP255        ---------------------N----------RR-K----------R--KV--------E-.DSK---I----------------H-------Q---R-----D-------R------E----N-I---R  129
F2.CM.MP257        -------------------K----N--------I-KR-A----------R---V--------EE.DSK---I----------------------Q---R-----G--------H------------I--QR  129
```

<- pol cds end

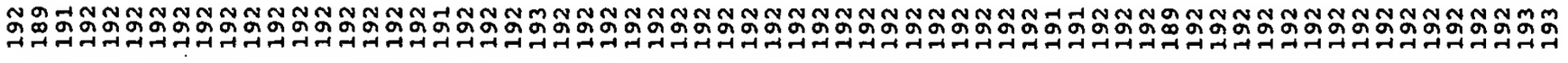
Hydrophilic region

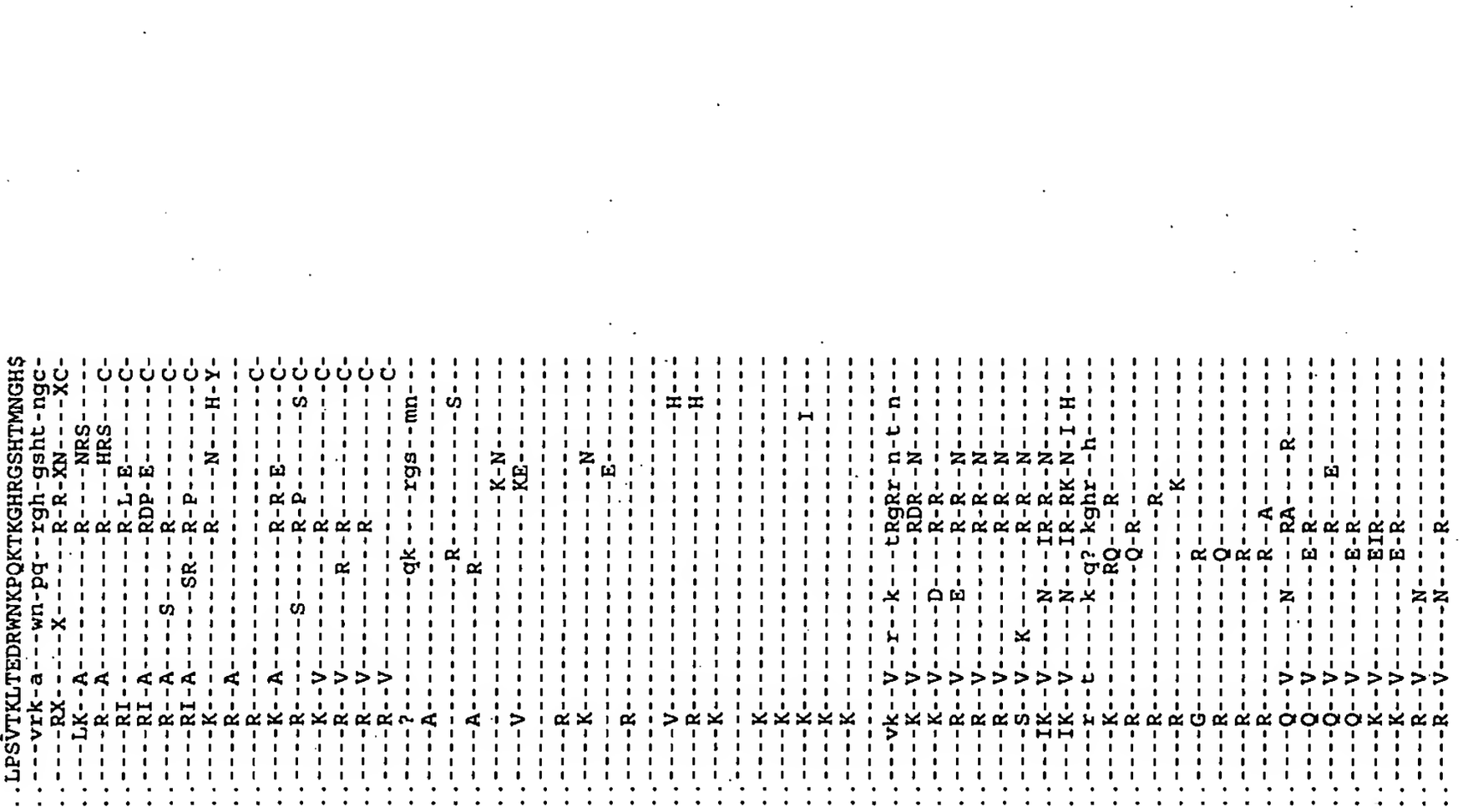
```
B.FR.HXB2                MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHYESPHPRISSEVHIPLGD..ARLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPELADQLIHLYYFDCFSDSAIRKALLGHI    128
G.BE.DRCBL               --------------------N--------I-K--K----------R--KV----------..-K--VR--------DKE----H-------QG--------D---H----H--N---E------I--QT    128
G.CG.CNG30               --------------------N--------I-K--K----------R---V---------N..-M-IVR---------K-----H-------Q---G--I--D--------H--K-------------Q-    128
G.FI.HH8793-1-1          --------------------N------L---K----$-----F--R--KV----------..-K--VS-----N---K--Q--H-------QR-----I--D--------H------E----N-I--R-    127
G.NG.92NG083             -------V------------N----------K--K----------R---V---------R-..-T--VR------A--K--Q--H-------Q------I--NT--H----------E------I--E-    128
G.SE.SE6165              ---------------------H----------K--------P--A-R---V----------..-T--V----------K--Q--H-------QR--R---E-D---H----H-------------I--Q-    128
H.BE.VI991               ------------------K--N---------K--KK-V-------TN-KT-------V--..-----------------------H-------QE-----I--D--------H-------------I---R    128
H.BE.VI997               ----------------NK------Y---I-K--NR-V-----D----------------..-----------------------H-------Q---------G-------TH------E----G-I--RV    128
H.CF.90CF056             ----------------N-------Y--HI-R----------F--T------------E..-----------N----E------------L-------E-G-------MH------E------I--RV    128
J.SE.SE7022              ----------------N-------Y--N--K---Q-L-----D-R--K----------E..----V------Q-----------------R---R-----G-------MH------------I--Q-    128
J.SE.SE7887              ----------------N-------Y--N--K---K-L-----D-N--K----------E..-I---------Q-----------------QR--R-----G-------MC------------I--Q-    128
K.CD.EQTB11C             ----------------N-------Y-----K--NR-R-----D-N--K----------..-E--V----------E------------L-K-R-----G-------I-------E----------R    128
K.CM.MP535               ----------------N-------Y--HI-K--NR-Y-------R--K----------..-E--V------L-----------------L---R---E-D-------I-------E--V---I---R    128
N.CM.YBF106              --------V-----X-K----N----------K--K--Y------TK--KT---X---V-X..-X---V-----T---QP----HX------QGK-K------M--K---C---N--TA----Q-V--RP    128
N.CM.YBF30               --------V-------K--K-N----------K--K--Y------TH--K--------V-Q..----TV-----T---QS----H-------LRK-K------M--K----H-----TA----Q-V--RP    128
CONSENSUS O              ---r--vl-v--v--QkVKa-N----Y-K-rsrktr?-y----?-srN-rv--g-Y--vgg..a?ivv------MP--rdEhl-H----e-qykk-k--I---T--Rm--lh--t--Te--i-k-I--QR    125
O.CM.ANT70               -------L-------QKVKA-N----Y-K-R-R-TEN-W-------RN--V--S-Y--V-V..-HV-V------MP----E---H-------Y-K-K--I---T--RM---H--T--TA--V---I--QR    128
O.CM.MVP5180             -------L----I--QKVKA-N----Y-K-M-K--AN-R-------RN-KV--A-Y--VAE..-DI-V------MP---EE---H------QY-E-K--I---T--RM---H--T--TE------I--QR    128
O.FR.HIVY16019           ------A--------QKVKV-N----Y-K-R-R-TKE-C----F-YKN--V--G-Y--V-M..-WI-L------MP-----E---H------QY---K--I---T--RM------T--TE------I--QR    128
O.FR.HIVY16020           ---K---L-L-----QKVKV-N----Y-K-R-R--KD-Y-------TN--V--S-Y--V-P..-HVLV------MP--K-E-M-H------QC-K-T--I---T--RM---H--T--TE------I--QR    128
O.FR.HIVY16021           -------L-------QKVKV-N----Y-K-R-R--KE-H-------RN-----S-Y--T-P..-HI-V------MP----E-M-H------QY-E-T--I---T--RM---H--A--TE------I--QR    128
O.FR.HIVY16022           -------L-------QKVKA-N----Y-K-R-RQT-E-H-------RN--V--G-Y--V-G..-NV-V------MP----E---H------QY-K-K--I---T--RI--QH--T--T-------I--QR    128
O.FR.HIVY16023           -------L-------QKVKA-N----Y-K-RT--T-N-Y----F--RN--V--G-Y----M..-CV-V------MP----EQ--H------QY-K-T--I---T--RI---H--T--T-------I--QR    128
O.FR.HIVY16024           -------L-------QKVK--N----Y-K-R--RT-Q-Y----F-HRN-----S-Y--V-A..-Y--V------MP----E---H------QY-K----I---T--RI------T--TE----Q-I--QR    128
O.SN.99SE-MP1299         -------L-------Q-VK--N----Y-K-R---T-D-Y----F--RN--V--G-Y--V-G..PWI-V------MP----E---H-------Y-K-K--I---T--RM--I---A--TE------I--QR    128
O.SN.99SE-MP1300         -------L-------Q-VK--N----Y-K-R---T-D-Y----F--KN--V--G-Y--V-G..PWI-V------MP----E---H----D--YRK-T--I---T--RM--T---A--TE------I--QR    128
CONSENSUS 01             -----------q------rt-n--------i-K--kk-f--------Qh-KV-------lge..-r-v-r-----q---K--h--H-------Qrk-s--i--D---ql---q--d---d----k-i--Qv    128
01_AE.CF.90CF11697       ---------------------Y--------I-K--KN---------Q--KV-------I--..-----R----------K-----H-------QRK-------D---RQ---Q-------------M--QV    128
01_AE.CF.90CF402         ------------------A-N--------S-K--AK---------Q--KV----------...-I-R----------K-----H-------QRK$---I--D--------Q---------------Q-    127
01_AE.CF.90CF4071        ----------------K--N--------I-K--KK-V-------Q--KV---------E..-----R----------K-----H-------QEK----I--D---R----Q------E------I--QV    128
01_AE.DE.K08DE           ----------R----------N--------I-K--KK---------Q--KV----------..-----K-----Q---K-----H-------QR--R--I--D--------Q-------------I--QV    128
01_AE.TH.93TH057         ---------------------N--------I-K--KK---------Q--KV---------E..-----R-----Q---K--Q--H-------QRN----I--D---R----Q-----------R-I--QV    128
01_AE.TH.95TNIH047       ---------------------N---------K--KK---------QN-KV--------EE..-K-I-R-----Q---K--Q--H-------QRT----I--D--------H-------R-I--QV    128
01_AE.TH.CM240           ---------------------N--------I-K--KK---------Q--KV---------E..-----R-----Q---K-----H-------QRK----I--D---K----Q--G---------I--QV    128
01_AE.TH.TH022           ----------R----------N--------I-K--KN---------Q--KV---------E..-K---R-----Q---K--Q--H-------QRK----I--D--------Q-------------I--QV    128
CONSENSUS 02             ---------------------N----h---v-?--k?------y--r--kVs---------..-rlvvr---g-ht-e-d-h--H----e-kQrr----I--D--------h--d--s?----k-i--qv    125
02_AG.CM.97CMMP807       ---------------------N---------K--NK---------R--KV----------..-T---K----------E----H-------QR-----I--D--------H--N---E----Q-I--QV    128
02_AG.FR.DJ263           ---------------------N--------I-R--KN------F--R--KV----------..---IVR------A--------H------KQRK----I--D--------H-----AE------I--EV    128
02_AG.FR.DJ264           ---------------------N--------I-R--KN------F--R--KV----------..---IVR---R--A-----Y--H------KQRK----I--D--------H-----T-------I--QV    128
02_AG.GH.G829            ---------------------N---------R--KD---------R--KV----------..--I-VR---------------H-------Q------I--D--------H--T---E------I--EV    128
02_AG.NG.IBNG            ---------------------N---------K--K-------F--R--KVC---------..----VR---------------H------KQ------I--D---------N----------I--E-    128
02_AG.SE.SE7812          ---------------------N----Y---K-R--KD---------S---V----------..----VR---------------H------KQR-----I--D--------H------------I--Q-    128
02_AG.SN.98SEMP1211      ---------------------N---------K--HR---------R--KV----------..-T-IVR-----QP--------H----V-QQ------I--D--------H------E---------QV    128
02_AG.SN.98SEMP1213      ---------------------N--------I-K--K----------Q--KV----------..----VR--------D---Q--H------KQRK----I--D--------H-------------I--QV    128
03_AB.RU.KAL153-2        ---------------------N------I-I-K-----V-K-----RN-------------..-K---K-----------------A------E--------N---------------E----N-I---R    128
03_AB.RU.RU98001         ---------------------N------I-I-K-----V-K-----RN-------------..-K---K-----------------A------E--------N---------------N-I---R    128
04_cpx.CY.94CY032-3      -A--------------K----N----------K--K-$---------R--KV----------E..-----VR-----QP--Q-----H-------LR--------D-------MH------E------I---R    127
04_cpx.GR.97PVCH         ----------------K----N--------I-K--K--S-------R---V----Q-----..------R-----QP--K-----H---M---LR--------D-------MH------E----Q-I--YR    128
04_cpx.GR.97PVMY         --------T-------K----N-------HI-K--K--V-K-----RN--A----------E..-K--VR-----QP-KK-----H----X--LRS-------D-------VH------E------I---R    128
05_DF.BE.VI1310          ------------------N-------Y--H--K--NR-C----F--RN-------------..-K--V-------------------------R--------S-------V-------E----N-I--R-    128
05_DF.BE.VI961           ------------------N-------Y--H--K-TKA---------R------------EE..-K---I----------E------------G-----I--G-------M-------E------I--YR    128
06_cpx.AU.BFP90          ------------------N-------Y--HI-K--KR-N-----D-N--K----------S..-E---------N----K------------LR--R-----SM------IH------E----------R    128
06_cpx.ML.95ML127        ------------------N-------Y--HI-K--KR-T-----D-N--KV---------C..-E---------N----K------------LR--R-----G-------MH------E------I--Q-    128
06_cpx.ML.95ML84         ------------------NA------Y--N--K--K--L-----D-N--K----------C..-E--V------N----E------------L-K-R-----S-------MH------E------I---V    128
06_cpx.SN.97SE1078       ------------------N-------Y--HI-K---K-A-----D-H--K----------K..-E--V------N----K------------L---R-----S-------MH------TE----E-I-----    128
10_CD.BFL061             -------L-------------N----------R--N--S-------R--K----------E..-S-I-K----------E----H-------L------I--G-------MH-----AE------I--QV    128
10_CD.BFL071             -------L-------------N--------I-RR-K----------R--K----------..---IVK----------E----N-------LR-S------G-------MH-----A-------I--Q-    128
10_CD.BFL110             -------L-------------N----------R-----Y-------R--K----------E..-----VK------P--KE----H-A-----LR-------G-------MH-----AE------IV-Q-    128
11_cpx.CM.97CM-MP81      ---------------------H----------K---R-M-------R--K----------..-D--V----------E------------M---R-----G-------IH------E----E-I---R    128
11_cpx.FR.99FR-MP12      ---------------------H--------I-K--KR-L-------R--K----I----R-..-E--V------------------------L---R-----G-------MH-----T-----T-I--YR    128
11_cpx.FR.99FR-MP13      ---------------------H-------HI-K-TT--L-------R--K----I-------..-K--V------Q----E----H-------L---R-----G---H---MH-----AE------I--QR    128
11_cpx.GR.GR17           -------------A-------N----------K--K--L-------R---V----------EE..DM--V----------E------------R---R-------------MH-----AE------I---R    128
CPZ.CD.CPZANT            I-K-$KIEMI--------NI--------IWETKVLKP-K------ND--KKGE------PTLDKK--V-VF---QC---P----H-------CGK-I------T---M--Q---P----Q-V-Q-I--ER    129
CPZ.CM.CAM3              -----L-----------VKA-N----C---KTK--KD---------S--KV---I---IEK..-K--V------MP-QHA----H-------CGS--------M--K----Q-----AA------V--KQ    128
CPZ.CM.CAM5              --------V---------K--N----T---RTK--KK----------DN-KV--------DT..-K--V------T--QHA----H-------CGS-K------T--R----H-----AA------V--KQ    128
CPZ.GA.CPZGAB            ------------------K--N----Y-I-R-K----------DH-N-KVA--I---FR-.YSK-IV----A-SP---A----H----Q--LGS-V-----FT--R---SQ-----AET---R-I--QL    129
CPZ.US.CPZUS             -------L---------VKA-N----Y---R-K--K----------TN--VN------IDK..-KI--I-----K---QA----H-------QGT-K--I---M--K----------TA------V--RP    128
```

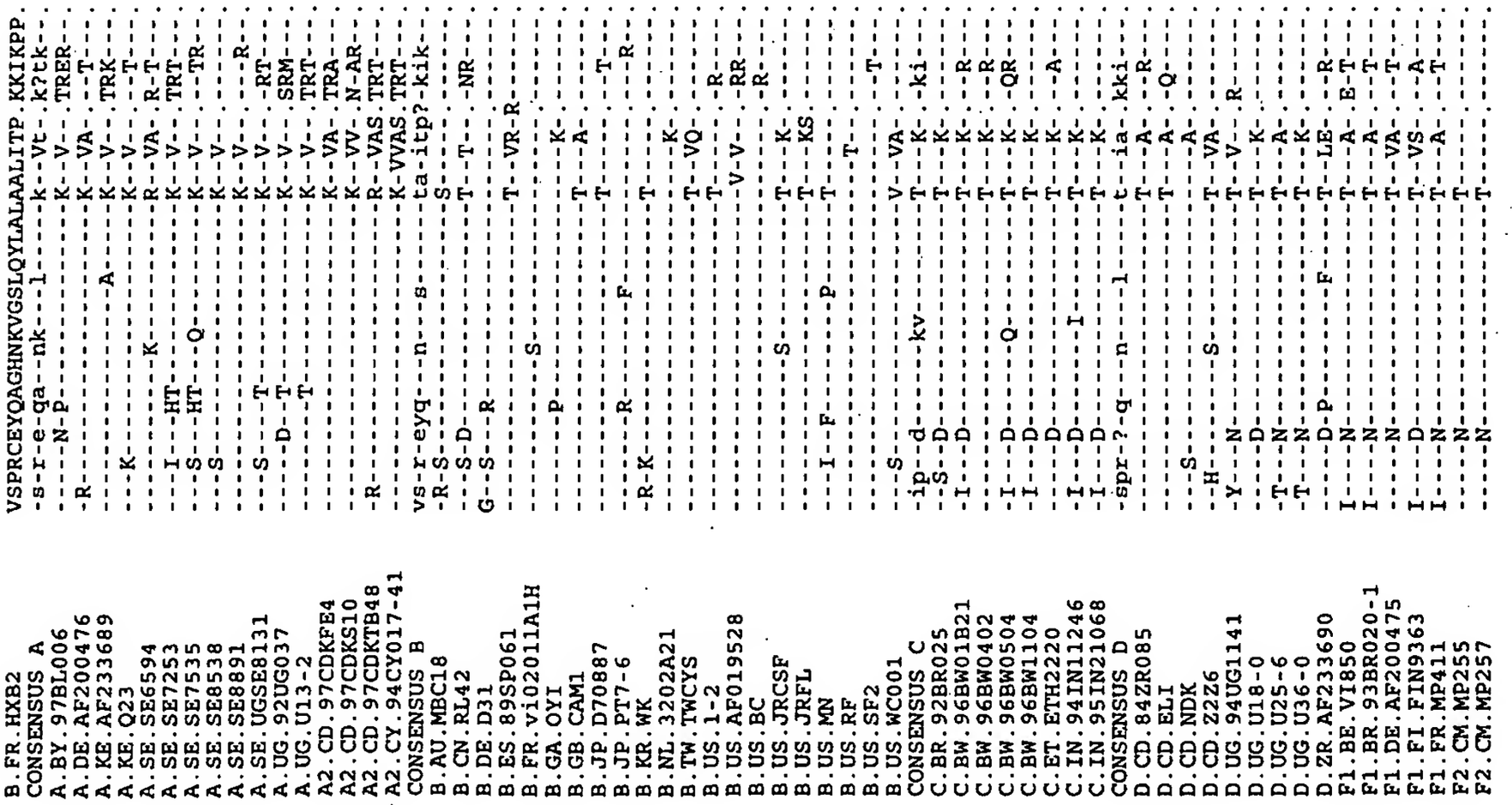
HIV-1/SIVcpz proteins

HIV-1/SIVcpz protein alignment: VIF

HIV-1/SIVcpz proteins







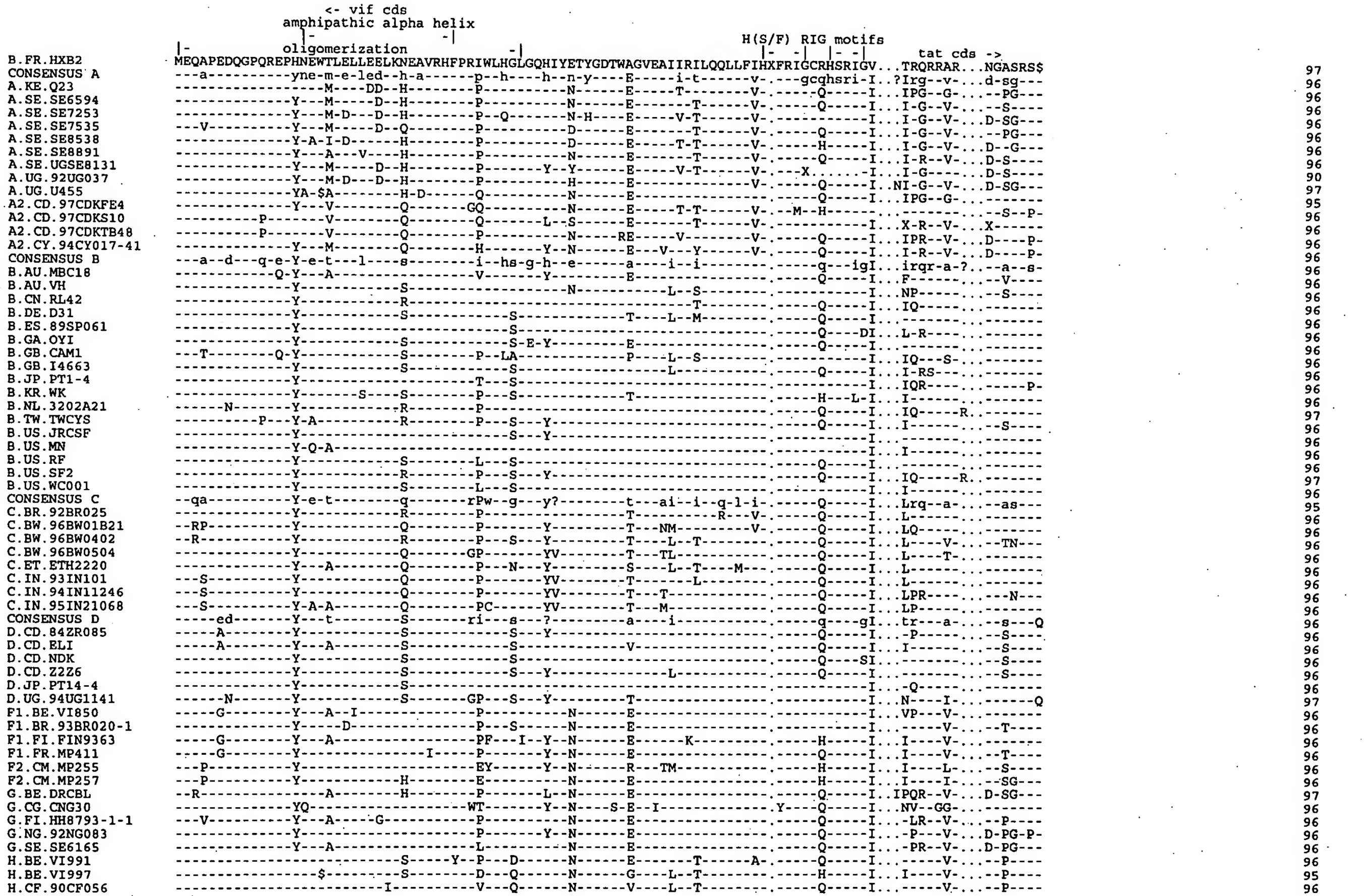
HIV-1/SIVcpz protein alignment: VIF

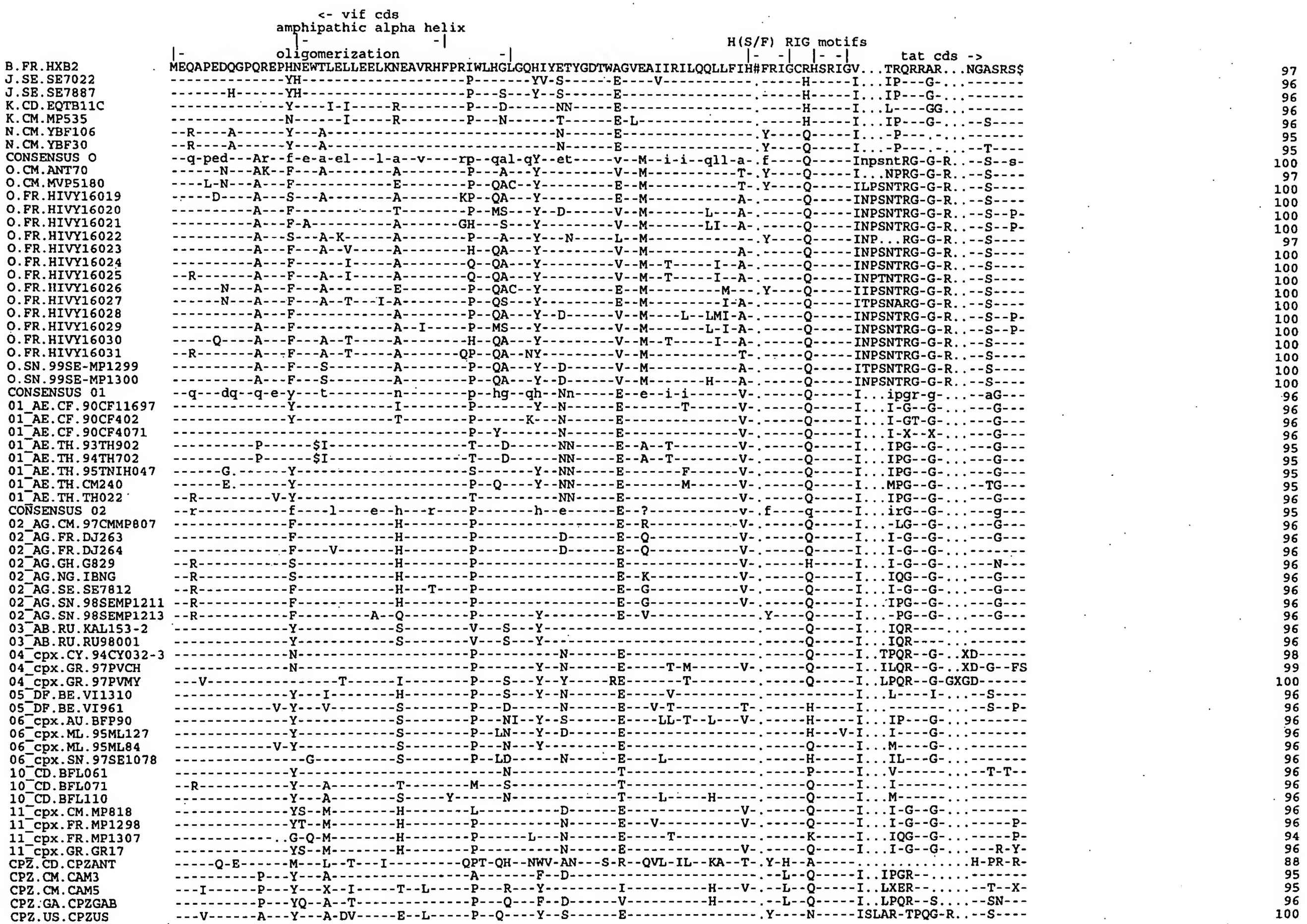
```
                                                                vpr cds ->
B.FR.HXB2              VSPRCEYQAGHNKVGSLQYLALAALITP.KKIKPP...LPSVTKLTEDRWNKPQKTKGHRGSHTMNGH$   192
G.BE.DRCBL             -R-S---P--------------KV-VA-.TRRR--...----R---------E----R---ENP------   192
G.CG.CNG30             -R-S---PT---Q---------K--V--.PR----...----R-------------R-----R-R---W   193
G.FI.HH8793-1-1        -R--------------------K--V--.T-TR--...----R---------S---R---E-P------   191
G.NG.92NG083           -------P-------------SK--V--.TRKR--...----G--A----------RD--ENP------   192
G.SE.SE6165            ------------Q---------KV-V-S.-RSR--...-----E-A----------R---ENP------   192
H.BE.VI991             ---I-D-----R----------T---S-.-RT----...----R--V----------R------------   192
H.BE.VI997             ------------Q---------T--VA-.--T----...----K--V--G-------R---------R--   192
H.CF.90CF056           -R---N-P---KQ--T------T--VA-.-------...----R--V----------R------------   192
J.SE.SE7022            -----D----------------T---K-.-RR----...----Q--V----------RD--E--------   192
J.SE.SE7887            -----D----------------T---R-.-RR----...----Q--V----------T---E--------   192
K.CD.EQTB11C           -----------TQ---------T---A-.--T-----...V---Q--V----------R--------S-Q-   192
K.CM.MP535             ------C----------------T--VA-.RRP-----...V---K--V----------R-----Q------   192
N.CM.YBF106            -L---X-P---XQ--T-----XT-XVGV.--RR----...-------------ER---Q----NPI---P-   192
N.CM.YBF30             -L-----P---KQ--T------T-WVGA.--R-----...-------------EH--MQ----NPI-----   192
CONSENSUS O            vLtk---pa--SQ--T--l---ravvKa.krnk----...----Qk----r-n?hlrIrDQle-hs-n---   188
O.CM.ANT70             -LTK---PT--SQ--T--L---R-VVKA.RSR-----...----Q---------HLRIRDQLK-PS-----   192
O.CM.MVP5180           -LTK---L---SQ--T--F---K-VVKV.-RN-----...----QR--------W-IRDQL---S-----   192
O.FR.HIVY16019         ILVK---P---SQ--T--L---R-IEKA.-RNR----...----Q-----K-S-HLMIRDQL---P-----   192
O.FR.HIVY16020         ILTK---S---SQ--T--L---R-VVKE.-RN-----...----Q---------HLRIRDQLE--S-----   192
O.FR.HIVY16021         -LTK---S---SQ--T--L---R-VVKE.-RN-----...----Q---------HLRI-DQLE--S-H---   192
O.FR.HIVY16022         -LTK---P---SQ--T--L---R-VVKA.-RS-----...----Q--------RHLRIRDQPE--P-----   192
O.FR.HIVY16023         -LTK---T---SQ--T--L---R-VVKA.-RN-----...----Q--------RHLRIRDQLE--S-----   192
O.FR.HIVY16024         ILSK---P---SQ--T--L---KVVVKA.RRN-----...----Q-----K-SRRLRIRDQLE--S-----   192
O.SN.99SE-MP1299       -LT----P---SQ--T--L---RVVVKE.-RN-----...----Q-------SRHLRIRDQLE--S-----   192
O.SN.99SE-MP1300       -LT----S---SQ--T--L---RVVVKE.-RH-----...----Q-------SRHLRIRDQL---S-----   192
CONSENSUS 01           -Rr--e-Ps-------------K--ttp.-rir----...----rkl-----nkp--?rg--EnP--n---   191
01_AE.CF.90CF11697     -R-----PT-------------K--A--.--TR----...----R-------------R---ENP------   192
01_AE.CF.90CF402       -R-----P--------------K--TKT.--T-----...----R----------------E-P------   191
01_AE.CF.90CF4071      -R---D-PE-------------K--A--.---R----...----R-------------R---ENP--S---   192
01_AE.DE.K08DE         -RH--D-PS-------------K--TA-.-R-R----...----K-------S----IRD--ENP------   192
01_AE.TH.93TH057       -RR----PS-------------K--T--.-R-R----...----K------------IW---ENP------   192
01_AE.TH.95TNIH047     -RR----PS-------------K--T--.-R-R----...----K---------H--.GD--ENP------   191
01_AE.TH.CM240         -RR----PS-------------K--T--.-R-R----...-----.EI----------.R---ENP------   190
01_AE.TH.TH022         -RR----PS-------------K--T--.-R-----...----R--------E---IR---EYP------   192
CONSENSUS 02           -rpr---q---nk---------k--Vt-.trtk----...----k--a----ne--k-r--r-sRs---hW   190
02_AG.CM.97CMMP807     -R--------------------K--VI-.TRTR----...----K----------R----K--RS----W   193
02_AG.FR.DJ263         -R---------KQ---------K--V--.A-T-----...----K--A----------R----NRS-----   192
02_AG.FR.DJ264         ------------Q---------K--V--.V-T-----...----K--A----------R----NRS---Q-   192
02_AG.GH.G829          -R--------------------K--V--.TRK-----...----R--A-----E----R-----RP---R-   192
02_AG.NG.IBNG          -R--------------------N--VA-.T-T-----...----R--A----KE----R-----RP-----   192
02_AG.SE.SE7812        -R---------T----------K--V--.TRT-----...----K--A-----E----R-----RS-----   192
02_AG.SN.98SEMP1211    -R-K------------------K--V--.TRR-----...----K--A-----E----R----NRS-----   192
02_AG.SN.98SEMP1213    -RS----P--------------K--V--.IRR-----...----K--------E----R-----RS---R-   192
03_AB.RU.KAL153-2      ---S---R----------------R--.-------...----------------R--D-------S---   192
03_AB.RU.RU98001       ---S---R----------------R--.-------...----------------R--D-------S---   192
04_cpx.CY.94CY032-3    -------------------------S-.--T-----...----K--V----------R-R-ENQI-----   191
04_cpx.GR.97PVCH       ------------E------------S-.--T-----...----K--V----------R-R-ENQI-----   192
04_cpx.GR.97PVMY       -------------P------------S-.--T-----...----K--V------S-----R-E--I-----   192
05_DF.BE.VI1310        ------F----------------T------.--T-----...----R----------------R--N-----Y-   192
05_DF.BE.VI961         ------------------------T------.Q-----...----R----------R------C-----Y-   192
06_cpx.AU.BFP90        -----D----------------T---K-.E-R-----...----Q--V----------R---E--------   192
06_cpx.ML.95ML127      ---Q-D----------------T---K-.--R-----...----Q--V----------RD--E--------   192
06_cpx.ML.95ML84       ---K-D----------------T---K-.R-R-----...----Q--V----------RD--EC-------   192
06_cpx.SN.97SE1078     -----DFE--------------K--VKT.-RR-----...----Q--V----------D--E--I-D---   192
10_CD.BFL061           -T---N-----KQ---------T------.--T-----...----R----------R--------------   192
10_CD.BFL071           -T---NF-----Q---------T------.--TR----...----R--------E--R--------------   192
10_CD.BFL110           -T---N------Q---------T------.--T-----...----R----------R-----E--I-----   192
11_cpx.CM.97CM-MP81    ------------Q---------K--V--.TRA-----...----R--A----------R----N--A--C-   192
11_cpx.FR.99FR-MP12    ----------------------K--VA-.TRT-----...----R--A----------R----N--L--C-   192
11_cpx.FR.99FR-MP13    -------P--------------R--V--.TRAR----...----R----------------..AI--C-   190
11_cpx.GR.GR17         -------P--R----------K--V--.-RT-----...----R------------R-R--N--V--C-   192
CPZ.CD.CPZANT          ILTY-H-KK--SQ--T-----FCKILEFRGYP-G-RRQF--LSI---------RRMR---ENQ$-S-C-   196
CPZ.CM.CAM3            -Y-K---P---QQ---------R-WVRV.G-K-----...-------------RH--IRV-QE-------   192
CPZ.CM.CAM5            -Y-K---PT--QQX--------R-WVGV.Q-KT----...---IX--A------Y---R--QEN----X--   192
CPZ.GA.CPZGAB          -A-----KE--RQ-----F---K---SE.RRHR----...----A---------H-R--V-QENL-R----   193
CPZ.US.CPZUS           -I-K---PT---------F---K--VGQ.S-RR----...--------------C-----QPEN---S---   192
```


HIV-1/SIVcpz proteins

HIV-1/SIVcpz protein alignment: VPR

HIV-1/SIVcpz proteins


<- vif cds
amphipathic alpha helix
oligomerization
H(S/F) RIG motifs
tat cds ->
B.FR.HXB2 MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGDTWAGVEAIIRILQQLLFIHXFRIGCRHSRIGV...TRQRRAR...NGASRS$
CONSENSUS A
A.KE.Q23
A.SE.SE6594
A.SE.SE7253
A.SE.SE7535
A.SE.SE8538
A.SE.SE8891
A.SE.UGSE8131
A.UG.92UG037
A.UG.U455
A2.CD.97CDKFE4
A2.CD.97CDKS10
A2.CD.97CDKTB48
A2.CY.94CY017-41
CONSENSUS B
B.AU.MBC18
B.AU.VH
B.CN.RL42
B.DE.D31
B.ES.89SP061
B.GA.OYI
B.GB.CAM1
B.GB.I4663
B.JP.PT1-4
B.KR.WK
B.NL.3202A21
B.TW.TWCYS
B.US.JRCSF
B.US.MN
B.US.RF
B.US.SF2
B.US.WC001
CONSENSUS C
C.BR.92BR025
C.BW.96BW01B21
C.BW.96BW0402
C.BW.96BW0504
C.ET.ETH2220
C.IN.93IN101
C.IN.94IN11246
C.IN.95IN21068
CONSENSUS D
D.CD.84ZR085
D.CD.ELI
D.CD.NDK
D.CD.Z2Z6
D.JP.PT14-4
D.UG.94UG1141
F1.BE.VI850
F1.BR.93BR020-1
F1.FI.FIN9363
F1.FR.MP411
F2.CM.MP255
F2.CM.MP257
G.BE.DRCBL
G.CG.CNG30
G.FI.HH8793-1-1
G.NG.92NG083
G.SE.SE6165
H.BE.VI991
H.BE.VI997
H.CF.90CF056
97
96
96
96
96
96
96
96
96
90
97
95
96
96
96
96
96
96
96
96



HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

intramolecular disulfide bonding | rev cds | nls | exon \ / exon

```
B.FR.HXB2          MEPVD...PRLEPWKHPGSQPKTACT.NCYCKKCCFHCQVCFI.TKALGISYGRKK..RRQRRRAHQNSQTHQASLSKQ.PTS.QPR.GD.PTGPKE$KKKVERETETDPFD$  100
CONSENSUS A        -d---...-nl---n----q-t-p-n.kcy--k--y--?v--l.n-G------r--..-kq--gtP-snKD--npipk-.piP.qtQ.g?.st-pe-Sk--v-SKaet-Rf?-   98
A.KE.93.Q23        -D---...-N----N------T---N.K-------Y------L.N-G---------..------G-P-S-KD--N-IQ--.-IP.-TQ.-V.S-----S---M-SKA-A-R---  101
A.SE.SE6594        -D---...-N----N------S---N.K----I--Y--P---L.N-G---------..-TR--G-P-SNKD--NPI----.-IP.-TQ.---.S---E-S-----SKA---R---  101
A.SE.SE7253        -D---...-N----N------T-P-S.K-------Y--I---Q.N-G---------..-------TP-S-KD--NPIP--.-IP.-TQ.-V.S---E-S-----SKA-A-R---  101
A.SE.SE7535        -D---...-N----N------T---N.K-------Y---A--L.N-G---------..-K---G-P-SNKD--NPIQQ-.SIP.-TQ.-I.----E-S-----SKA---R---  101
A.SE.SE8538        -------...-NI---N------T-P--.K-------Y--PA--L.N-G---------..-K---GTP-GNKD--NPIP--.S-P.-TQ.-I.S---E-S-----SK--A-R---  101
A.SE.SE8891        -D---...-N----N------A---S.A-------Y--P---L.N-G---------..------TP-S-KD--NPVP--.SIP.-AQ.---.S---E-S-----SK----R-A-  101
A.SE.UGSE8131      -D---...-N----N------T-P-N.KYF--R--Y--L---Q.H-G---------..------GTP-S-KD--DPIP--.SIP.PAQ.-I.----E-S-----SKA---R-A-  101
A.UG.U455          ------...-N----------T---S.-----V--W---L--L.K-G------K--..-KP--GPP-G-KD--TLIP--.-LP.-SQ.RV.SA-QE-S-----SKAK--R-A-  101
A.UG.UG273A        -D---...-N----N----K-T-P-S.K----V--Y--P---L.K-G---------..-S---GTP-SNKD--NPIP--.-IP.-TQ.-I.S-----SQ----SK----R-A-  101
A.UG.UG275A        -D---...-N----N------T-P-N.K-------Y--T---L.N-G---------..-K----TP-CNKD--NPIP--.-IP.-TQ.-V.S---E-S-----SK----R-T-  101
A.UG.92UG037       -D---...-S----N--------P-N.K----V--Y---C--L.N-G---------..-KP--GTP-SNKD--NPIP--.-IP.RTQ.---.S---E-S-----SK--A-RYA-  101
A2.CD.97CDKS10     ------...-K----N----------N.K-------Y---L--L.N-G---------..--P--GPS-S-KD--NPIP--.SLP.-TQ.RV.S--TE-S--E--SK--P-R---  101
A2.CD.97CDKTB48    ------...-N----N----------N.--------Y---R--L.N-G---------..--P--GPD-GNTD--NPVP--.SLP.-TQ.R-.S---E-S-----SKA---R---  101
A2.CY.94CY017.41   ------...-K----N----------.K----R--Y---L----.N-G---------..--P--KPSPSNKD--NPIP--.SLP.-AQ.RV.----E-P--E--SKA---R---  101
CONSENSUS B        -e---...-r----k------k-act.n----k--l------t.t-g---sy----..------raPqdsqt--vs-s--.pas.qpr.--.pt-pk-skk---r--etd-vdQ  101
B.AU.MBC200        ------...-----------------.T----------.--G---------..-------P-D-E---V-----.---.-H-.--.------P----------V--  102
B.CN.RL42          ------...-----------------N.--------L------M.K-G---F-----..-------P-D-----------.-A-.---.--.-A----S-----S------R--  101
B.DE.D31           ------...------E----------.----------T.K-G---------..-------PED-----V------.-A-.---.--.------S-----T------I--  101
B.ES.89SP061       ------...-------------R----.----------.--G---------..-------P-D-----V-----.RA-.--G.--.------S----------VN-  101
B.GA.OYI           ------...----------SN.-----R--L------T.K-G---------..-------P-D-K---V------.-A-.---.--.------S----------E--  101
B.GB.CAM1          ------...-------------R----.-----Q----------.--G---------..-------TP-S-K----------.-A-.-FQ.--.------S-----G----H-G--  101
B.KR.WK            ------...---------------P--.K-------L------M.--G---------..-------P-DNKN--V------.---.RA-.--.---QE-S-E---K--VV--VT-  101
B.NL.3202A21       ------...-K-------------N.----------------T.K-G---------..------SP-D-E---V------.-A-.---.--.------S----------V--  101
B.TW.TWCYS         -D---...----------------.--------L------M.--------N----..------TP--------D----.---.---.--.Q--Q--ST----------N-Q  102
B.US.WC001         ------...-S----------------.-----Y--L------T.R-G---------..-------P-D-----V------.-A-.---.--.------SQ----S------SA-  101
B.US.RF            ------...----------------N.--------Y------L.--G---------..------GPP-G-----V------.---.---.--.------S-E----------AVQ  102
B.US.MN            ------...----------------.T----------T.K----------..-------PED-----V--P---.-AP.-F-.--.------S----------H-V--  101
B.US.JRCSF         -------...-S----------------.--------L------T.--G---------..-------PP-D-----V--P---.-S-.-Q-.--.------S----------DN-  101
B.US.JRFL          -------...-S----------------.T-------L--------.--G---------..-------P-D----------.-S-.---.--.------S----------D--  101
CONSENSUS C        me-v-...-n----n------k---n.?-y--k-sY--L---q.tkG-g-s------..------saPpS-ed--nlisk-.-Lp.rTq.-d.p--sE-S-----sKtet-pfd-  100
C.BR.92BR025       ------...-N----N----------N.------R-SY--L---Q.--G---------..------S-PPS-ED--NPIP---.-LP.-T-.---.Q--SE-S-----SK--------  101
C.BW.96BW01B22     ------...-N----N----------N.-----H-SY--L---Q.--G---Y------..------S-PPS-ED--NLI----.-LP.RTQ.--.---SE-S-----SK-K------  101
C.BW.96BW0402      ---I-...-N----N----------N.K-F--H-SY--L---Q.--G-S--------..------S-PPS-ED--NPV----.-LP.-T-.--.---LE-S-----SK--------  101
C.BW.96BW0504      ------...-N-----------R-----.K----Y--Y--L---Q.--G---------..------STPPS-ES--NLI-E-.-LP.RTQ.-N.---SE-S-----SK--A---A-  101
C.DJ.DJ259A        SV$I-...-------N----------N.K------SY--L---Q.--G---------..------S-PPS-ED--TPI----.-LP.-T-.--.----E-S-----SK--A-R---  100
C.DJ.DJ373A        --------...-N----N----------.K------SY--L---Q.--G---------..------S-PSS-ED--NLI----.-L-.RTQ.--.----E-S-----SK-KA--L--  101
C.ET.ETH2220       -------...-N----N----------N.Q------SY--L---L.-XG---------..--------P-S-KD--NLI----.-L-.HT-.--.---SE-S-----SKA----YA-  101
C.IN.93IN101       -------...-N----N----------N.Q----R-SY--L---Q.K-G---------..------S-PPS-ED--NLI----.-LP.RTQ.--.---SE-S-----SK-K------  101
C.IN.94IN11246     ---I-...-N----N----------N.-----H-SY--L---Q.R-G---------..------S-P-S-ED--DLI----.-LP.RTQ.--.---SE-S-----GK-TS-----  101
C.IN.95IN21068     -------...-N----N----------N.-------SY--L---Q.K-G---------..------S-P-S-ED--NPI----.-LP.RTP.--.---SE-S-----SK-K------  101
CONSENSUS D        md-v-...-nlep-N------r-p-N.k-y--k--Y--qv--i.t-G---------..--q--rppqgg-ah-dpiPk-.-S-.-p-.-d.---p-eQ-----Sea-tdp--W  102
D.CD.ELI           -D---...-N----N------R-P-N.K-H-----Y--P---L.N-G---------..------GPP-GG-A--VPIP---.-S-.---.--.------Q-----S-A----$   99
D.CD.NDK           -D---...-N--S-N------R---N.K-H-----Y--------.--G---------..------KPP-GD-A--VPIPE-.-S-.-S-.--.------K------S-A------W  101
D.CD.84ZR085       -D---...-NID--N------R---N.--------Y--------.--G---------..-------PPHS-----DPIP---.-S-.---.--.---Q--------SK--VH---W  101
D.CD.Z2Z6          -D---...-NI---N----------N.R-H-----Y--------.--G---------..-------PS-GG----DPIP---.-S-.---.--.------------S-A------W  101
D.SN.SE365A2       -D---...-N----N--------P-N.--------Y---N---.--G---------..-------TP-GN----V-VP---.-S-.-H-.--.------------SKA---Q--W  101
D.UG.UG274A2       SV$I-...------N------S-P-N.K-------Y---L---.--G---------..-------PP-S--A--DPIP---.-S-.----.-N.-----K------SKA-A----W  100
D.UG.94UG1141      -------...-N----N------R-P-N.K----R--Y--Y---V.--G---------..--P---TPPGG-AN-DPVP---.-S-.----.-N.------------S---A----W  101
CONSENSUS F1       --l--...-n-D--Nh-----t-P-t.k----r-----?w--t.--G---------..-r--h-TP-S--i-qd?vPk-.-?-.-a-.gn.-----eS--E--SKakt--c?-   97
F1.BE.VI850        ------...-S-D--N------T-P--.K----R------W--T.--G---------..----H-TP-S--V--N--P---.-L-.-A-.--.------S--E--SKAK---CA-  101
F1.BR.BZ126        --L-...-N-D--N------T-P--.K----R------W--T.--G---------..-K--H-TP-S--I--DLVP---.-I-.-A-.-N.------S--E--SKATR--CA-  101
F1.BR.BZ163        --L-...-N-D--N------T-P--.K----R------W--T.--G---------..-K--H-TP-S--I--DLVP---.-I-.-A-.-N.-----KS--E--SKA----$   99
F1.BR.93BR020.1    --L-...-N-D--N------T-P--.R----W-----YW--T.--G---------..----P-TP-S--I--DFVP---.-I-.-A-.-N.------S--E--SKAK---$   99
F1.FI.FIN9363      --L-...-N-D--N------P-P-N.K----R-----YW--A.--G---------..----H-TP-S--I--DPVP---.-L-.---.RN.------S--E--SKAK---C--  101
F1.FR.MP411        --L-...-N-D--NQ-----T-P--.K----------YC--A.--G---------..------TP-S--S-KNPIPE-.-L-.-A-.--.------S--E--SK--------  101
F2.CM.MP255        --V-...-KID--N------E-P-N.----------PL--M.K-G---------..------TP-G-KI--DPVP---.-L-.-T-.--.----E-S-----SQ-----$   99
F2.CM.MP257        --V-...-N-D---------E-P-N.K----------L--T.R-G---------..------TP-SGEV--DPV----.-L-.-T-.---.-K--E-S-----SK-K---S--  101
```

```
                                                intramolecular
                                               disulfide bonding
                                               |-              -|  rev cds ->|-  nls  -|          exon \ / exon
B.FR.HXB2            MEPVD...PRLEPWKHPGSQPKTACT.NCYCKKCCFHCQVCFI.TKALGISYGRKK..RRQRRRAHQNSQTHQASLSKQ.PTS.QPR.GD.PTGPKE$KKKVERETETDPFD$   100
G.BE.DRCBL           -D---...-K-------------P-N.-----SVAAL.----L.N-G---------..-KH--GTPHS-KD--TPVP--.-F-.TT-.-N.----Q-S--E--SK--------   100
G.FI.HH8793-12.1     -D---...-N----N----------N.K----I--W------L.N-G---------..-KH---PP-DNKD--NPVP--.-LP.TT-.-N.------S-----SK-K------   101
G.NG.92NG083         -D---...-K----N------T-P-N.K----V--W------L.N-G---------..--P--GTP-G-KD--NPVP--.-LP.ITS.-N.---SEKP--E-ASK-----L--   101
G.SE.SE6165          -D---...-N----N--------P-N.K-F--V--W------L.N-G---------..-KH--GTP-S-KG--DPVP--.-LP.TT-.-N.------S--E-ASKA-A-QC--   101
H.BE.VI991           -D---...-NQ---N------R---N.--------Y---L--L.K-G---Y-----..-----GTPKSL-D--TLIP--.-L-.RTS.--.----EK-----ASK-------W   101
H.BE.VI997           -D---...-NQ---N---N----P-N.-----Q-S----L--L.K-G---------..-SR--ATPASV-D--NHIP--.-L-.RT-.--.----------ASK--A--C--   100
H.CF.90CF056         -D---...-K----N------Q---N.--------Y---M--L.K-G---------..-S--H-TPASL-D--N-I---.-L-.RTH.--.------Q--E-ASK-----$    99
J.SE.SE7887          ------...-NR---N----------.--------Y------L.Q-G---------..-----S-PPG-K---DLIP--.-L-..TQ.RK.----E-S--E--SKA-P-R---   101
J.SE.SE7022          ------...-NR---N----------.--------Y------L.Q-G---------..-----S-PPG-KN--DLIPE-.-LF.-TQ.RK.----E-S--E--SKA-P-R---   101
K.CD.EQTB11C         ------...-NI---NQ---------N.Q-------Y---L--L.Q-G---C---E-..----TTTPYA-KN-KDPIP--.-LP.-A-.--.------S--E--SK-K---$    99
K.CM.MP535           -D---...-NI---NQ---------N.Q----R--Y---I--L.K-G----N----..--P--TTPY--EN--DP-R--.-L-.----.-E.Q-D---S-----SK-K--Q---   101
N.CM.YBF106          ------...------N----------N.K-------Y--MC--T.K-G---------..-S----PPKS-KD--DPIPE-.-L-R-QP.--.Q--Q-KQ--AL-GK--A--C--   102
N.CM.YBF30           ------...------N----------N.-----R--Y--LY--T.K-G---------..-S----TP-S-KS--DLIPE-.-L-..QQ.--.Q--Q-KQ-EAL-SK--A--C--   101
O.CM.ANT70           -D---...-EVP--H------QIP-N.-----R--Y--Y---V.R-G---------..-GRP..-AASHPD-KDPVP--.SPT.ITK.RK.QERQE-QEEE--KKAGPGGYPR   100
O.CM.MVP5180         -D---...-EMP--H----K-Q-P-N.-----R--Y--Y---T.K-G----H----..--RP.A-AASYPDNKDPVPE-.SL-.HTG.RK.QKRQE-QE----K--GPSGQPC   101
O.SN.MP1299          -D---...-EMP--H------Q-P-N.K----A--Y--Y---A.S-G---------..-.R-PA-AARHPDN-DIVPE-.L-Y.ITN.RK.QKRQE-QE-E--N-ACPR$    97
O.SN.MP1300          -D---...-EMP--H------Q-P-N.K----A--Y--Y---A.S-G--L------..-.R-PA-AARNPDN-DIVPE-.-PPITNN.RK.HKRQE-QE-E--K--GS-RYPR   102
10_CD.BFL061         ----H...-N----N----------N.K-H--A-SY--L---Q.---X--------..-----S-PPS-ED--NLI---.-S-.---.--.-------------A-A----W   101
10_CD.BFL071         ------...-N----N--------P-N.K-F--A-SY--L---Q..-G---------..-----S-PPS-ED--NLI---.-S-.---.--.-----------SQA-------   100
10_CD.BFL110         ------...-K----N--------P--.E----H-SY--L---Q.K-G---------..-----S-PPS-ED--DLI---.-S-.---.--.-----------SKA-A----W   101
CONSENSUS 01         --l----.-n----n------t---s.K----k--W---L--L.K-G----y-----..-KH--gtpqssKd--npIpk--.-lp.ii-?gn.--d---S--e-asKae--pcdW   102
01_AE.CF.90CF11697   -------.-S----N------T---S.K-------W---L--L.K-G---------..-KH--GPPPG-KD--NPIP--.-LP.TT-.-N.------S--E-AKKAK----A-   101
01_AE.CF.90CF402     -------.-N----------T---S.K----M--W---L--L.K-G---------..-KH--GPS-D-KD--N-IP--.-LP.TS-.-N.------S-----SKA----$    101
01_AE.CF.90CF4071    --L----.-K----N------T---N.K----I--W---L--L.K-G---------..-KH--GTP-G-KG--DPI---.-LP.II-.-N.------S--E--SK-K---CA-   101
01_AE.TH.CM240       --L----.-N----N------T---S.K-------W---L--L.K-G---------..-KH--GTP-S-KD--NPIP--.-LP.II-.RN.--D---S--E-ASKA---QC--   101
01_AE.TH.93TH057     --L----.-N----------S---S.K----I--W---L--L.K-G---------..-KH--GTP-SRKD--HPIPE-.-L-.II-.-N.--D---S--E-ASKA----C--   101
01_AE.TH.93TH065     ------...-N----N------T---S.K----I--W---L--L.K-G----H----..-KH--GTP-S-KD--SPIPE-.-LP.II-.-N.--D---S--E-ASKA----C--   101
01_AE.TH.93TH253     --L----.-N----N------T---S.K-------W---L--L.K-G----H----..-KH--GTP-SRKD--YPIPE-.-LP.II-G-N.--D---S--E-ASKA----C--   102
01_AE.TH.95TNIH047   --L----.-N----N------T---S.K-------W---L--L.K-G----H----..-KH--GTP-S-KD--TPIR--.-P-.II-.-N.--D---S-----SKA----$    100
01_AE.TH.TH022       --L----.-N----N------T---S.K-------W---L--L.K-G----H----..-KH---TP-S-KD--YPIPE-.-SP.II-.-I.--D---S--E-ASKA---QC--   101
02_AG.CM.MP807       -------.-S----N------T---S.K----M--W---L--L.N-G---------..--L--GTP-SR-D--N-V----.-LP.TT-.-N.-----E-S--E-ASK-----$    99
CONSENSUS 02         --lv----.-s----N----q-T---S.k----i--w---L--L.N-G---------..-rR--GTP-SrqDh-nPVp--.pLP.tT-.-n.ptg---S-ke-asKte-dpcd-   101
02_AG.FR.DJ263       ------...-S----N------T---S.K----I--W---L--L.N-G---------..--R--GTP-SR-D--NPVP--.-LP.TT-.--.--D---S--E-ASK----QG--   101
02_AG.FR.DJ264       ------...-S----N------T---S.-----I--W---L--L.N-G---------..-XR--GTP-SR-DR-NPV---.-LP.TT-.-N.------S-RE--SK-K---C--   101
02_AG.GH.G829        --LI-...-S----N------T---S.K----I--W---L--L.N-G---------..--R--GTP-SH-D--NPV---.SLP.-T-.--.------S--E--SK------A-   101
02_AG.NG.IBNG        --L----.-S----N------T---S.K----M--X---L--L.N-G---------..--R--GTP-SR-D--NPVP--.-LP.TT-.-N.--D---S--E--SK-K---C--   101
02_AG.SE.SE7812      --L----.-S----N------T---S.K----I--W---L--L.N-G---------..-KR--GTP-SR-DN-DPVP--.-LP.TT-.-N.-A----S--E-AGK-----C--   101
02_AG.SN.MP1211      --L----.-K----N------T---S.K----L--W---L--L.N-G---------..--R--GTP-SR-D--NPVP--.-LP.TT-.-E.Q-----S--E-ASK---G-C--   101
02_AG.SN.MP1213      --L----.-S----N----K-T---S.K-------W---L--L.N-G---------..--R--GTP-SRPD--NPVP--.-LP.NT-.-N.------S----ASKAK---$    99
03_AB.RU.KAL153-2    ------...-----------------.--------L------M.K-G---------..-------P-DN--D-V--P--.-A-.---.--.----------M------H----   100
03_AB.RU.RU98001     ------...-----------------.--------L------M.K-G---------..------PP-DN--D-V--P--.-A-.---.--.----------------H----   100
04_cpx.CY.94CY032-3  ------...-D----N------T-D-N.K-F-----W------L.K-G---------..-KH--GSL-G-KG--NLIP--.-L-Q--N.--.S---E-Q----ASK--A---A-   102
04_cpx.GR.97PVCH     --A--LVY-N----N------T---N.K-F-----W---A--L.K-G---------..-KRNEDLLGF-RDR-NPIP--X-L-.--N.-N.-E----Q----ASK--A-----   105
04_cpx.GR.97PVMY     -------.-N----N------T-----.K-F--V--Y------L.R-G---------..-KH---PP-G-RDR-NPIP--.-L-Q-HS.--.------Q-EA-ASK--SN----   102
05_DF.BE.VI1310      -D---...-N----N------R---N.Q----M--Y---N---.--G---------..--P---PP-G--A--DPVPE-.-P-.---.--.-----KQ--E--SK--A-Q--W   102
05_DF.BE.VI961       -D---...-N----N------R---N.Q-H-----Y--------.--G---------..--S---PP-GG-A--IPVPE-.-S-.-A-.--.---Q--Q-----SKA------C   102
06_cpx.AU.BFP90      -------.-KI---NQ---R-------.K-------Y--P---L.N-G---------..-----Q-PPG-KN--DPV---.-L-.-TQ.RE.Q---EKS--E--SKA-P-R---   101
06_cpx.ML.95ML127    -------.-NI---NQ---R-------.P------SY---L--L.N-G---------..-----T-PPG-KN--DPVP--.-L-..TQ.RG.----EKS-----SKA-P-R---   101
06_cpx.ML.95ML84     -------.-KI---NQ---R-------.S----R--Y--PL--L.N-G---------..-----T-PPG-KN--DPVP--.-L-..TQ.RK.S---E-S--E--SKA---R---   101
06_cpx.SN.97SE1078   -------.-NI---NQ---R-------.S----Q--Y--P---L.N-G---------..-----T-PLG-KS--DPVP--.-L-..TQ.RE.S---EKS-EE--SK--P-R---   101
11_cpx.CM.MP818      -------.-NI---NQ---R------N.Q----R--Y--PH--L.K-G---Y-----..-----A-SHS-EN--DPIP--.-ST.--N.RK.----E-S--E--SKA-P-R---   101
11_cpx.FR.MP1298     -------.-KI---N----------N.Q-F--R--Y---I--L.K-G---------..-----T-YHD-KN--DP-PE-.-LP.-T-.RK.----E-S--E--SK--A-R---   101
11_cpx.FR.MP1307     ------...-NI---N------Q---N.Q----R-SY---I--L.K-G---------..-----ATPYS-KN--DPVPE-.-LP.-TT.RK.--D-E-S--EM-SKA-PTR---   101
11_cpx.GR.GR17       ---G-...INID--NQ---------N.Q----R--Y---H--L.K-G---------.W----T-SRS-KN--DPIPE-.-LP.-AS.RN.----E-P--E--SKA-PA----   101
CPZ.CM.CAM3          -------.-D----L------R---N.-----M-----I---E.K-G---------RRG-R-LS-ASSTH.N-G--QQ-.-L-.RQQ.-N.Q--QE-Q--TL-SQAKA---A-   102
CPZ.CD.CPZANT        -D---...AETP--L--PAT-A-P-N.-----C--Y--PL--T.K-G--------RRA--N--TTAES-ENN-DPV--$.SLP.KTS.RI.QSSQ-K-E----EK-GSGGGPC   102
CPZ.CM.CAM5          -------.XN----L----------N.-----C-----M---E.K-G---------RRG-RHSL-SSSTH.N-D-IRQ-.-L-.RQQ.-N.Q--EE-Q--T--SQA-A--R-W   103
CPZ.GA.CPZGAB        -D-I-...-D----------R-V-N.-----A--Y--IY--T.K-G---------.RTTR--T-PAG-KNN-D-IP--.-L-.-S-.-N.KE-SEKST-E-ASK--A-Q$    101
CPZ.US.CPZUS         -D---...-NI----Q---------N.----------V---T.K-G---------.RG-R--TTAKS-ADN-DP-R--.-F-..QL.-N.Q--Q--Q--AL-KQ-TS-QR--   102
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

high affinity binding site — Leu-rich effector domain

exon \/ exon — nls

```
B.FR.HXB2          MAGRS.GDS..DEE.LIR.TVRLIKLLYQSNPPP.NPEG.TRQARRNRRRRWRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQ.......GVGSPQILVESPTVLESGTKE$  116
CONSENSUS A        --g--.-ns..-ee.-Lk.Ai-i--il-q---y-.k-k-.s---rK-r----rA--r--dsl-e-i-stC--rsa---p-q--pl-r-hl-csed-g----qqsqg?etgvgrp-vs?-s?vv-g---kn-  116
A.KE.Q23           -----.---.----.-L-.A--I--I--K---Y-.K-K-.S----K-------A-----D-------S-C---PT----------H---C---------QSQGAET---RH-VS----VI-G----N-  120
A.SE.SE6594        -----.-PG..----.-LK.AI-I--I------Y-.K-R-.-----K-------A-----D-------SAC----T----------N---R---------QSQGVET---R--VSG---VI-G----N-  123
A.SE.SE7253        -----.---.----.-L-.A--I--I------Y-.R-K-.S---QK-------A-----D-------S-C----E---------S--T-H---HD--------QSQGVET---R--VPG-PS---G----T-  123
A.SE.SE7535        -----.-N-..----.-L-.AI-I--I--N---Y-.K-K-.S----K-------A-----D-------NAC--G-T----------N---R---------QSQGIET---R--VS----VI-G------  123
A.SE.SE8538        -----.-N-..----.-L-.AI-I--I------H-.K-R-.S----K-------A-----D-------S-C----------G-H---C---------E.........---R--VS----G--D----N-  116
A.SE.SE8891        -----.---.----.-LK.A--I--I------Y-.K-R-.-----K-------A--K--D-L-----S-C---PQ-----L---I---H---S---------QSQGAET---G--VSE--SII-G----T-  123
A.SE.UGSE8131      -----.---.----.-LK.A--T--I------Y-.Q-K-.S----K-------A-----D-L-Q---SAC----E----------HI--H---------G...........T-VSG--SV--D----D-  112
A.UA.ukr970063     -----.-NT..-A-.-LT.A--I--FV-----Y-.T-R-.-----K-Q-----A-----N-L-----SAC---------S-----I---H--YS---------QSQGTET---G---SG--SGI-G------  123
A.UG.92UG037       -----.-NP..----.-L-.AI-I--I------Y-.E-K-.-----K-------A-----DTL---V-S-C---P----------I---H---S---------QSQGVET---RT-VSG---V--G----N-  123
A.UG.U455          --R---.-NP..--D.-LK.A--I---------C-.--R-.S----K-------A-----D-L-----SDC---P----------I---R---S-S-------QPQGTET---G---S---SA--G----N-  123
A.UG.UG273A        -----.-V-..----.-LK.AI-I--I------Y-.K-K-.S----K-------A-----D-L-Q---S-C---P----------H---R---E-----.........---RN-VS--P-V--G---QN-  116
A.UG.UG275A        -----.-S-..----.-LN.AI-I--I------Y-.K-K-.S----K------KA-----D-L-----S-C---P----------I---H---S---------G...........---VS---TSI-G----N-  112
A2.CD.97CDKS10     -----.--P..--D.-L-.A-KI--I------Y-.K-R-.S----K-------A--N--D---------C---PT---------I---H---S---------QSQGTET---G---FG-PS---G--IEK-  123
A2.CD.97CDKTB48    -----.--P..--D.---.AI-I--I------Y-.K-R-.-----K-------A-----D-------S-C---PT----------H---S-----C---QSQGAET---R--TS---SGI-G--IED-  123
A2.CY.94CY017-41   T----.D-P..--S.-LQ.AI-T--I------Y-.K-R-.S---Q-------A-----D-----V-R-C---PT----------HI--S--------LQSQGTET---RS-ES---SVI-G---E--  123
CONSENSUS B        ----s.g--..dee.-Lk.tv-l--fl----pp-.spe-.-rq----------e--rqirsi-gw--sty--rpaep-p-q---le--t--csed--t--tq.......v-s-qil----av-es-tkee  117
B.AU.MBC18         -----.---.----.-LK.------SI--------.----------------------L----A---S-F---PE-S----------------------R.......M----------P---T--T-E  117
B.AU.VH            -----.---.-D-.-LK.------F---------.S---.------------A-----RQ--GW--S-----P----------------SK--------.....---G--V-----A-----AA--  116
B.CN.RL42          -----.E--.----.-LK.-------------L-.S---.------------A-----RE--D---V----G-T----------------SK--------.....---G--V-----A-----AA--  116
B.DE.D31           -----.---.----.-LK.------F---------.S---.------------Q-----Q-------S-----PEQ----P----------S---------.....----------A--D------  116
B.DE.HAN           -----.---.----.-LK.------F---------.SN--.-PT----------------R-------S-F---P----------------S----N----.....------V-----A---P-----  116
B.ES.89SP061       -----.---.----.-LK.------F-----A--.SQ--.----------------------R---GW--SN----P----------------S---------.....----------I-D------  116
B.GA.OYI           -----.---.----.-LK.------F---------.----.----------------------RK--GW--S----------------N---S---------.....----------E------A---P-----  116
B.GB.CAM1          -----.---.----.-LK.A----------L-.SSK-.----------H-RA--NW--S-H---P----------------SK--------.....----------A---------  116
B.KR.WK            -----.---.----.-L-.-I-I--F------L-.E---.------------R--WW-Q-L-GW--N-H---P----------------E-------.....---N--------A---------  116
B.NL.3202A21       -----.---.--D.-LK.------F---------.S---.----------------R-------S----------------D---------.....----------A---------  116
B.TW.TWCYS         -----.---.----.-L-.-------I--------.----.----------RT--GW--SN----P----------------D---------.....-------F--------D------  116
B.UA.AF193278      -----.---.----.-LK.-I----F---------.----.----------HL--------S-----PE-----P----------S---------.....----------D-----X  117
B.US.JRCSF         -----.---.--D.-LK.------F---------.SN--.----------RT------S-----P----------------------.....---N-E----------------  116
B.US.JRFL          -----.---.----.-LK.----------------.S---.----------RT------S-----P----------------------.....----------E----------  116
B.US.MN            -----.---.----.-LK.------F---------.SS--.----------H-R---AW--SN----P----------Q------S---------.....----------.E  115
B.US.RF            ----R.---.--D.-LK.A-----S---------.S---.----------RRC--W--D------VD--Q----------SS--------.....------V-----A-----A---  116
B.US.SC            -----.---.E--.-LK.------F---------.K---.----------R---GW--SN----L----------------------.....----------------------  116
B.US.SF2           -----.---.----.-L-.----------------.S---.----------R---GW--S----------------S----N--A-.....----------A--D------  116
B.US.WC001         -----.---.----.-LK.------F---------.----.----------A-----RAL-GW--SN----P----------I-------S---------.....----------A--------  116
CONSENSUS C        -----.g-s..-ea.-Lq.Avri-ki---s--Y-.k-e-.--q-rk-------A--r--hsise---stcl--pa---plq---i---hidcSesggts--qqsqgttegv-s-$isGkscav-gsgakKE  123
C.BR.92BR025       -----.---.--A.-LQ.A--I--RI------Y-.K---.----------A-------------S-CV--P-----F----I---NIN-S-SG------QPQGNTER--N-$VFGRPCA----RV-KE  123
C.BW.96BW01B21     -----.---.--A.-LQ.A--I--I------Y-.K---.----W--------A-------------S-C---PT---------I---HI--S-SS-A----QSQGTTE---N-$VSGK-FA--G--A-KE  123
C.BW.96BW0402      -----.---.--A.-LQ.A--I--I------Y-.K---.----W--------A-------------S-C---PT---------I---HIGDS-SG------QSQGTTE---N-$VSGK-CAI-G-RA-KE  123
C.BW.96BW0504      -----.---.--A.-LQ.A-KA--I------Y-.E-K-.-----K-------A-------L-----SAC---P-----F----I---HI--SKSV-----E.........-----$-SGKPCA--G-RA-KE  116
C.BW.96BW1104      -----.--N..--A.-LQ.A--I--I------Y-.K---.-----K-------A--K--N-------S-C----------FL---I---HISDS-SG------QSQGTPE---N-$-SGK-CA--GARA-KE  123
C.BW.96BW1210      -----.---.--A.-LQ.A--I--I---N--Y-.K---.-----K-------A-----D---T---S-C---PE----F----I---NIGDS-SG------QSQGTTE-----$-SGK-CA--G----KE  123
C.BW.96BW15B03     -----.E--.--A.-LH.A--I--I------Y-.E-K-.-----K-------A-----D---T---S-C---PE----F----I---NIGDS-SG------QSQGTTE-----$VSGK-CA--G----KE  123
C.BW.96BW16B01     -----.----..AA.-LQ.A--I--I------Y-.K-K-.-----K-------A-----RA--A---SAC---P----------HIGSS-SG------.....-----$-SGKPCA--G-RA-KE  116
C.BW.96BW17A09     -----.--N..--A.-LQ.AMGI--I------Y-.K---.--R--K-------A-------------S-C---PT---H-P-------H---S---EPT--EQSQGATE-----$-SGKPWA--G--A-KE  123
C.DJ.DJ259A        -----.---.--A.-LQ.A--I-R------Y-.K---.------------A-----D-------SAC---P----------I---N---S-SG--A--E.........---N-$VSGKPCA--G--A-KE  116
C.DJ.DJ373A        -----.---.--A.-LQ.A--I--I------Y-.E-K-.----Q-------A-------------S-C---P-----F----I---N---S-SG-----E.........---N-$VSGKPCAI-G--A-KE  116
C.ET.ETH2220       -----.---.----.-LK.A--I--I------Y-.T---.------------A------TL-----SNF---P----------N---S--S------QSQGTTE---N-$-SGKPCA--G--A-KE  123
C.IN.93IN101       -----.---.--A.-LQ.A--I--I------Y-.E-K-.-----K-------A-------------S-C----T---------I---HIGGS-SG------.....QL---$-SGK-CA--G--A-KE  116
C.IN.94IN11246     -----.---.--A.-LK.A--I--I------Y-.E-R-.-----K-------A-------------S-C---P----------I---HISGS-SG------QSQGTTER----$-SGK-CA--G--A-KE  123
C.IN.95IN21068     -----.---.--A.-LK.A--I--I------Y-.E-R-.-----K-------A-------------S-F---P----------I---HISGS-SA-----PQSQGTTE-----$-SGK-C---G--AEKE  123
C.SO.SO145A        -----.---.--A.-L-.A--I--I------Y-.K---.-----K-------A-------------S-C---PT---S-----I---HIGGS-SG------QSQGTTER----$VSGK-WP--G--A-KE  123
C.UG.UG268A2       -----.---.--T.-LQ.A--I--I------Y-.----.------------A-----RE-NQ---S-C-------------------I--S-SS------QSQGTTE-----$VSGKPCA--G-R--KE  123
CONSENSUS D        -----.--s..-ed.-Lk.Avr---?---s-----.Sp--.------------A---q-hs-g---ls-y---se---------l--ln-N-s-d-gts---------.g-gsp-is--s-av-?sg-eeQ  115
D.CD.84ZR085       -----.---.--D.-LT.A-----I---------.S---.------------A---Y----G----S------E----------IN-N-S---------.....-------S----AI---R-E--  116
D.CD.ELI           -----.---.--D.-LK.A-----F---------.S---.------------A-----RE-A----------P----------N-N-S---R-----.....---H---S----------E-Q  117
D.CD.NDK           -----.---.--N.-LK.AI----F---------.S---.------------A--------G---IC-F---PE----------N-N-S---------.....-------P--P-A------E--  116
D.CD.Z2Z6          -----.--R..--D.-LK.A-----I---------.S---.------------A--------G----S------E----------N-N-S----A----.....-------S----------E-Q  117
D.SN.SE365A2       -----.---.----.-LK.AI----F---N----.ST--.------------A-----N--G----S----------------V---N-N-S---------.....R-----VS----A--DTE-EK-  116
D.UG.94UG1141      -----.--R..----.-LQ.A-----I---------.S---.------------A--------G---IS-----FE----------N-N-S---------.....----H--S----A--D------  116
D.UG.UG269A        -----.---.--D.-LK.A-----F---------.S---.------------A----------S-----PE----------N-N---N-------.....----H--S----KI-D------  116
D.UG.UG274A2       -----.---..-AD.-LK.A-K---I---------.S---.------------A--------G----S------E----------N-S----NP----.....-E-H--S----A--D------  116
CONSENSUS F1       -----.-ds..dt-.-L?.Avry--I------Y-.k-e-.------------A--rq-Rai-e---ssC---pe---------l---hINcSed-g.......QGaeE---sp-iSg-sh?--?------  113
F1.BE.VI850        -----.----..-T-.-LK.A-KC--I------Y-.K---.------------A-----RAL-D---SSC----E----------HIN-S----.......QGPEE----S--SG--HA---------  116
F1.BR.93BR020-1    -----.---..-Q-.-LK.A--Y--I------Y-.K---.------------A-----RE--D---SSC---P----------HIN-S----.......QGAEE----S--SG--H---G------  116
F1.BR.BZ126        -----.-N-..-T-.-L-.A--Y--I------Y-.K---.------------A---E-RAL-----SSC---PE----------NINSS---E.....QGAKE------TSG--HA--G------  116
F1.BR.BZ163        -----.-N-..GT-.-L-.AA-Y--I------Y-.K---.------------A-----RE--R---SSC---PE---------P----INNS-N-E.....QGAEE------TSG--HA--G------  116
F1.FI.FIN9363      -----.----..-T-.-LK.A-KY--I------Y-.S-D-.------------A-----RA------SSC---LE----------HIN-S----.......QGTEE-------SG-HH---------  116
F1.FR.MP411        -----.--N..----.-L-.A--A--I------Y-.K---.------------A--K--R-------VAC---PE----------HIN-SK---.....QGTNE---N---SM-PR---------  116
F2.CM.MP255        -----.---.----.-LK.A--Y--I------Y-.K---.------------A------E------T-----PT---LF---------IN-S----.....QGSER-------S---HAA-G--I---  116
F2.CM.MP257        -----.--R..----.-LK.A--Y--I------Y-.KL--.--K---------A------Q------S-C---LQ---R----L--K-HIN-S----.....QGTEK-------S---RA--G------  116
```

high affinity binding site Leu-rich effector domain
exon \/ exon |- nls -| |- -|

```
B.FR.HXB2          MAGRS.GDS..DEE.LIR.TVRLIKLLYQSNPPP.NPEG.TRQARRNRRRRWRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQ.....GVGSPQILVESPTVLESGTKE$  116
G.BE.DRCBL         -----.-ST..----.-LT.A--I---------S-.P---.-----------A----------S-C---PE----------H---S--G------QSQGTEI-------F---SV--G------  123
G.FI.HH8793-1-1    -----.-ST..--A.-L-.-I-I--I------Y-.P---.-----K-------A----------SSC----------------H--SS--GE---P-QSQGTET---G---F---SV--G--A---  123
G.NG.92NG083       -----.--P..---.-L-.A--I--T------Y-.S-A-.-----K-------A----------SAC---P-----F------G-S---SK-G------QPQGTET---R--V---P-V--G------  123
G.SE.SE6165        -----.-ST..---.-L-.A-KA--I------Y-.P---.-----------A-----SA------TA----P----------H---S--S------QPQGTET---RS$-F---SV--GQ-----  122
H.BE.VI991         -----.--N..--G.-L-.AC-I-R-------Y-.E-A-.----Q--------A--------G--V-A-C--GP----------S-------EK.....------TSG---A--GT-A---  116
H.BE.VI997         -----.-AG..--Q.-PQ.VCKI--II-----Y-.E-A-.----------A-----RA------TDC---PP----------K------EK.....-K-G---P---S---GT-----  116
H.CF.90CF056       -----.-A-..-T-.-LQ.VCKI--I------C-.E-T-.----------A-----RE------TSC---PP---T----------N-S-------EK.....-E-----SL--S-I-GT-----  116
J.SE.SE7022        -----.--N..-DQ.-LL.A--I--I------YS.K-N-.S----------A--N--D-------SSC---P----------I---R---S----N----.....---D---SG-PCM--GA-----  116
J.SE.SE7887        -----.---..-DQ.-LL.A-----I------Y-.K-N-.S----------A--N--D------PSSC---P----------I---R---S----N----.....---D---SG-PCM--GA-----  116
K.CD.EQTB11C       ----R.----..EQQ.-LT.P--I--I------Y-.K---.----------A-----RE--Q-V-SSC----T----------S-N-D--S-.....QGTEGEL-----P--PD---G--D---  116
K.CM.MP535         ----R.--P..--Q.-LT.---T--I--E---Y-.-L--.S--T---------A--K--S-----L-SAC----------I-K-N-N-D--P-.....KGTEG-L-----S--PC----------  116
N.CM.YBF106        -----.-VN..--D.-LK.A-KI--I------Y-DSSQ-.----------A-----RA------A---GPQ---D-P----G-----HK--S-DP--ESQQGTATTE$TQNT--GNNSI-GKRV-R-  123
N.CM.YBF30         -----.-VN..---.-L-.A--V--I------Y-..-SK-.----------A-----RA------SSC--GPP---D-P----D-----TE--S--P--ESQQGTATTE$TQNT--GNTCI-GKRV-G-  122
O.CM.ANT70         -----.E-...-Q..-LQ.AIQI--I------Q-.S-R-.S-N--K-------R--A-VDTLAA-V-A-VVHGPQNNNIVD-----Q-SIRDP-GDQL-EAWTVDPRAEDN$CL-N-CSCN-I-ATRIA--  120
O.CM.MVP5180       -----.EE...-QQ.-LQ.AIQI--I------C-.T-A-.S-N--K-------R--A-VD-LAT---A-VVHG-QDNNLVD-----Q-SIRDPDGDQP---WTVDSGTEDN$CL-T-HSCN-I-ATRVA--  120
O.SN.MP1299        -----.DG...-QP.-L-.AIQI--I-----SHT.S-T-.S-S----------T-HAHVDTLAA---A-VVHGPQDNNLVE-----Q-NIRDP-ADRLP--GTVDPGTKDN$-L$T-WSCNAI-ATRIEK-  120
O.SN.MP1300        -----.DG...-QP.-L-.AIQI--I------H-SPTT-.STS----------K--A--DTLAA---A-VVHGPQDNNLVE-----Q-SIRDPDGDQP---WTVDPGTEDN$CL-N-HSCN-I-ATRVA--  121
CONSENSUS 01       -----.-sT..--e.-Lr.avri-ki-yq---y-.ss--?t--t-k----r--a------rais----s?Cl---t---p-q---l---hl--s--c------QSqGTETg--r---Sg--svI-Gp--kN-  122
01_AE.CF.90CF11697 -----.-NT..--D.-LQ.A--I--I------Y-.P---.-----K-------R-------L-----VACV---T----------HI--S---------QSQGTET---G---SG--SVI-G----N-  122
01_AE.CF.90CF402   -----.-ST..--D.-L-.---I--I------Y-.PA--.-----K----K--A------K-G----S-C----P----------H---S---------QSQGTET---G---SG--SVI-G----N-  123
01_AE.CF.90CF4071  -----.-ST..----.-L-.AA-A--I-F----Y-.SS--.-----K-------A-----RAL-----SAC----------H---S---------QSRGTET---R---SG--SVI-G---EN-  123
01_AE.TH.93TH057   -----.-ST..----.-L-.A--I-NI-----Y-.SS--.---T-K-------A-----RA------I-C----T----------H---------QSQGTET---R---SG--SVI-GP---N-  123
01_AE.TH.93TH065   -----.-ST..----.-L-.A--I--V------Y-.SS--.---T-K-------A-----RA------NACV---T----------H---S---------QSQGTET---R---SG--SVI-G----N-  123
01_AE.TH.93TH253   -----.-ST..----.-L-.A--I-NI-----Y-.SS--G---T-K-------A-----RA------S-C----T----------H---S---------QSQGTET---R---SG--SVI-GP---N-  124
01_AE.TH.95TNIH047 -----.-ST..----.-L-.A--I-----E-----.SS--.---T-K-------A-----RE------SSCV---T-----P----------H---S---------QSQGTET---R---SG---VI-GP---N-  123
01_AE.TH.CM240     -----.-ST..----.-L-.A--I---I------F-.SS--.---T---------A-----SA------S-C----T---------V---N---S--G------QSQGTET---R---SG--SVI-GP---N-  123
01_AE.TH.TH022     -----.-ST..----.-L-.A-KI-NI-----L-.SS--.S--T-K-------A-----SA------SAC----T---S-P----------H---S---------QSQGTET---R---S---SGI-GP---N-  123
CONSENSUS 02       -----.gda..-ee.-L-.?v-iikI-----Y-.p---.t--a-k---r---A----irai-e-i-S-C---sa---plq---i---c---s-dc-t----qsqgtetgvgspqis?--sii-Gs-----  121
02_AG.CM.97CMMP807 -----.-----.-A-.-L-.A--I--I------Y-.P---.-----K-------A-----RE------S-C---PT---------I---C---C-G-------QSQGTET-------SG--SII-GP-----  123
02_AG.FR.DJ263     -----.--A..----.-L-.V--I--I------Y-.P---.---T-K-------A-----RA----F-S-C----------I---C---S-G-------QPQGTET-----P-SG--S-I-G------  123
02_AG.FR.DJ264     -----.X-A..----.-L-.V--IV-I------Y-.P---.-----K---G---A-----RA--Q---S-C----------C---S-G-------QSQGTET-L-----SG--SDI-GA-----  123
02_AG.GH.G829      -----.--A..----.-L-.VI-I--I------Y-.K---.----------A-------L-----S-C---PE---SF--------N---S----N----.....S-AD---PG--RAI-G------  116
02_AG.NG.IBNG      -----.--A..----.-L-.A--I--I------Y-.P---.-----K-------A-----RA------S-C----------I---N---S---------LSQGTET-------S---YII-G------  123
02_AG.SE.SE7812    -----.-NA..----.-L-.A--T--I------Y-.P---.-----K-------A-----RA------S-C----------I---R---S---------.....------S---SI--G------  116
02_AG.SN.MP1211    -----.--A..--G.-L-.A--I-RI------Y-.P---.S----------A----VRA------S-C---P------P---I---C---S--S------QSQGTET----S-TS---SVI-G------  123
02_AG.SN.MP1213    -----.--A..----.-L-.V--T--I------Y-.T---.-----K-------A------K------S-C----------I-------S---------.....-------S---SII-G------  116
03_AB.RU.KAL153-2  -----.---..----.-LK.-I----F---------.S---.----------H------Q--S-----PE---L-H----------S---------.....---------D------  116
03_AB.RU.RU98001   -----.---..----.-D.-LK.-I----F---------.----.----------H-----Q---S-----PE-----H----------S---------.....---------D------  116
04_cpx.CY.94CY032-3 -----.-NI..--D.-FK.AA-A--I------Y-N--T-.----------A--K----L-----A-----P----------K---N-S-------DK.....------VS--L-A--GT-A---  117
04_cpx.GR.97PVCH   -----.-SX..N-D.-LGXS-GIV-I------Y-.--T-.--K---------A--K--------V-A-----P----------K---N-S-------EK.....------VS--P-A--GT-A---  117
04_cpx.GR.97PVMY   -----.-ST..--D.-LK.A-GIV-I------Y-N-TA-.----------A--NR---------AAC---P----------I-K-----S-------DK.....---N--VP--P-A--GT-D---  117
05_DF.BE.VI1310    -----.--R..--D.-LK.A-----I------L-.S---.----------A-----N--G--L-S------E----------N-N-S---------.....-------S--P-AI--------  116
05_DF.BE.VI961     -----.--R..--D.-LK.A-----F---------.R---.----------A-----R--AD--VD-----PE----------N-N-S---------.....-------S---------A---  116
06_cpx.AU.BFP90    -----.---..-DR.-LL.A--I--I------Y-.K-N-.S----------A--N--D-----V-S-C----------I---R---S----N----.....---N---SG-PDML-GT--T--  116
06_cpx.ML.95ML127  -----.--N..--Q.-LL.A--I--I---N--Y-.K-N-.G----------A--N--D-------S-C---PT----F----I---R---S----N----.....---D---PG-PGV--GT-----  116
06_cpx.ML.95ML84   -----.--N..--Q.-LL.A--T--I------Y-.K-S-.S----------A--K--D-------S-C----------I---R---S----N----.....---D---P--PGVL-GT-----  116
06_cpx.SN.97SE1078 -----.---..--Q.-LW.A--V--I------Y-.KLS-.S----------A--N--D-------SSC----E---------I---R---T----N----.....---D---PG-PGVL-GT-----  116
10_CD.BFL061       -----.---..--A.-LQ.A--I--I---------.S---.----------GG---DA----PE----------N-N----------H.....---GH--S-----I-D------  116
10_CD.BFL071       -----.---..--A.-LQ.A--I--I---------.S---.----------A----------CDC---PE----------N-N----------.....-------S----AI-DP----Y  117
10_CD.BFL110       -----.---..--A.-LQ.AA-I--I------L-.S---.----------A--------G---IS------E----------N-N----------.....----N--S--P-EI-D------  116
11_cpx.CM.MP818    -----.--N..--Q.-LT.A-KI--I------Q-.--T-.S----------A--N--D---Q---SDC----E---------I---H---S----NP---.....---DS--SG--D---GPR-E--  116
11_cpx.FR.MP1298   -----.--N..--Q.--T.---I--I------Y-.K---.N----------K--A--K--D---Q---S-C----E---S-P---I---H-H-S--S-D----.....---NS-TSG--S---G-R-E--  116
11_cpx.FR.MP1307   -----.---..-AQ.-LT.A--I--I------Y-.K-Q-.S--TQ--------A--N-LD---Q---S-C---PE----F----I---H--SS----D----.....---DS--P---N-I-GPR-E--  116
11_cpx.GR.GR17     -----.---..-AQ.-LA.AA-I--I----S-Y-.K-A-.--------Q-----A--N-L----Q---S-C----E-----P---I---H---S----N----.....---ES--PG--S---GPR-E--  116
CPZ.CM.CAM3        -----E--D..-AC.-LQ.A--I--V--S.--Y-.DNK-.----------A-------L-----SS-----P--PA-L---I----I-TG--I-KP--EPQQGTATAEELQ--P-G-DNL-GK-ATN-  123
CPZ.CD.CPZANT      ----EELEGT.--Q.-LK.A-KI--I------Y-.K-A-.S-A----K----KK--D-VEGLAA-V-R-LV-GPD-HN-VD--D-QN-S-EPLGSPP-PLQN.....YFQTV-PSG-QSQHPLDNNQN-  118
CPZ.CM.CAM5        -----E--E..-TR.-LQ.A--I--I--D.--Y-.DNK-.----------A------E-GQ-V-SSC----S--PD-L---ID----NPV--SAN---EHQQGTATVE-L$Y-PKG-TNI-GKRDTN-  122
CPZ.GA.CPZGAB      -----.EPQ..-DAR-LQ.A-KI--I------Y-.S---.--K---------A--K--SE--G-V-A-----PPK-GD-E--E-DK-S-Q-V-TTQDV--SNTSQPQTAT-ETVPAGGNYSI-GK-A-N-  124
CPZ.US.CPZUS       -----G--A..--Q.-L-.A--I--I--A---F-.-NS-.----------KN-----NE--Q---S-C----S--PD-P---IGG--INSE--STNP--EHQQGTATVE$TQST-LGGNRL-GK-A-S-  123
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins


```
                                                signal peptide \/ gp120
                                                      <- vpu cds end
                                                      *                                *                *                ^^^                                                *
B.FR.HXB2                MRVKE....KYQ.HL.WRWG..WRWGTMLLGMLMICSAT..EKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCV  120
CONSENSUS A              -rvmGiqr.Ncq.hl.wr--.....i-ilgmiIi-saa..en----------v-kd-e----------dt-v----------n---ihle----e-n----n--e--ht----------k---  119
A.GB.MA246               --AMGIQR.NC-.S-.---.....I-I---II----A..GN----------D-E----------T--------------------IN-E----K------N------T----------Q---  119
A.KE.K89                 ---MGTQR.NC-.--.L---.....I---II----A..-N----------N-E----------E--K----------------IH-E----E------N------T----------Q---  119
A.KE.Q23                 ---MGIQR.NC-.--.LT--.....I-I--TIIF---V..-N----------RD-D----------E--K----------------IH-D----K------N------T----------Q---  119
A.RW.PVPI                ---MGIQM.NC-.N-.---.....-I----I---VA..GN----------D-E----------A----------------IL-K----E------G------T----------Q---  118
A.SE.UGSE8131            ---MGTQR.NC-.--.LN--.....I-I---II---TA..-N----------D-E----------EK---------------LY-E----D------N------T----------Q---  119
A.UA.ukr970063           -KARGMQR.N--.--.---.....---FW-IIM-K-A..-N----------RD-E----------K----------------IS-E----K-D----N-----QT----------Q---  119
A.UG.92UG037             ---MGIER.N-P.CW.-T--.....I-I---II--NTA..-N----------I--D-N----------------S---LKME----E------N------T----------Q---  119
A.UG.U455                ---MGIQR.N-P.C-.---.....--I--LII--N-...QQ----------D-V----------A----------------ID------E------N--D----------Q---  118
A2.CD.97CDKS10           ---MGTQT.S--.--.---.....ILI----I--K---.D.W----------D-E--------D--E--A----------N-K----D------N----------Q---  118
A2.CD.97CDKTB48          T--MGTQR.NC-.KW.-E--.....ILVF--I-M-K-A...D----------RD-D----------A--K----------N-A----D------N-----A----------Q---  118
A2.CY.94CY017-41         ---MGTQR.N--.--.--G-.....ILI----IM-K--...D----------D-D-I----------IN-E----------N-----Q----------Q---  118
CONSENSUS B              -rvkgiRk.nyq.hl.-rw-??????tml--mlm-csaa..enl----------k-at-------d-kaydt-v-----th----t--n---vvle----n-n-w--n-ve--h-di-s---q------  119
B.AU.MBC18               -T--GTRK.N-R.--.-T--.....---I--I------A..NNF----------R----------E----------E-I----------N----------Q---  119
B.CN.RL42                ---TGIRK.N--.--.---.....----------N-A..-N----------G----------N----------Q---  119
B.DE.D31                 -K---IRK.N--.--.---.....----------..----------K----------D------E----D------N----------Q---  119
B.DE.HAN                 -K---TRK.N--.R-.--G-.....-L-------S-VA..GN----------MG----------Q----------  119
B.ES.89SP061             ---T-IRK.N--.--.-K--.....----------K-A..-N----------E-K----------N--D----------  119
B.GA.OYI                 -TARGTRK.N--.R-.---.....----------A..-N----------R--A----------G------D----N-----Q----------  119
B.GB.CAM1                --A-GIRK.NC-.R-.---.....----------A..D----------E----------N----------  119
B.GB.WB                  ----GIRN.N-...-.K--.....----------K-A..-Q----------D----A-E----------N----------  117
B.JP.ETR                 -----IRK.N--.--.----IML----------A..-Q----------A----------N-------N----------  125
B.JP.JH32                ----GIRK.N--.--.---.....-----I------A..-Q----------A----------E----K------N----------  119
B.KR.WK                  ----GIRK.N--.-W.---.....I-----W-----A..----------I-------F-E----------N----------  119
B.NL.3202A21             -K---TRK.N--.--.---.....----------A..-Q----------S----------A-E----D-D----N----------N----------  119
B.TH.93TH067             -K-TGIRK.N--.--.---.....----------A..-N----------N----------G----------N----------  119
B.TT.QZ4589              ---RGIRR.NC-.--.-T--.....-------I------..-Q----------K----------S------G----------N----------  119
B.TW.TWCYS               ---RGTRM.NC-.--.---.....---------S--A..-N----------T----------T-G----------N-AD--Q--------E------  119
B.UA.UKR1216             -----IRK.....--.---.....-LF----------..-NS----------SK--------Y-------S---IP-K------S-G--N----------  116
B.US.DH123               ---MGIRK.N--.--.-KG-.....-L---I------A..-Q----------N----------IL-E----D------N----------  119
B.US.JRCSF               ----GIRK.N--.--.-KG-.....IL---T------V..----------T----------E----D------N-----Q--V-N----------  119
B.US.JRFL                ----GIRK.S--.Y-.-KG-.....-L---I------V..----------E----H------N-----Q----------  119
B.US.MNCG                ----GIRR.N--.-W.-G--.....-----L-------..----------Q----------E----------N----------  119
B.US.RF                  ---M-MRK.NC-.--.-K--.....----------A..-D----------E----K--------K----------L-E----------N----------  119
B.US.SF2                 -K--GTRR.N--.--.---.....-L----------..----------R----------G----------N-----Q----------  119
CONSENSUS C              -rvmGiqr.Ncq.qw.-iw-.....ilgfwmlmicngm..gnl----------ke-kt--f----ak--et-v----a-h-------n---mvle----n----k-d--d---q-i-----qs-----  119
C.BI.BU910112            --ATGIQR.NC-.QW.-I--.....ILGFW-----NVM..GN----------A----------M--K----D---------D---Q----------  119
C.BR.92BR025             ---EGIQR.NWK.QW.-I--.....ILGFW-V--YNVR..GN----------K----------A----------M--E---------E--------Q----------  119
C.BW.96BW0402            ---MGIQR.NC-.QW.-I--.....ILGFW-IINGMGS.....----------KA--------R---R----------N-E----------D----------  116
C.CN.AF268277            ---TGIRK.N--.--.---.....----------A..-N----------K----------EK---------------M--E----------N-----V----------  119
C.DJ.DJ373A              ---MGIQM.NC-.QW.-IG-.....ILGFW--LM-NGM..GN----------Q--NP----------E----------S---L-ME----------G--D---Q--------EG-----  119
C.ET.ETH2220             -K-MGIQR.NC-.QW.-I--.....ILGFW-----NGM..GN----------D-SP----------G-F-------S---LG-E----------Q--------G-----  119
C.IN.93IN101             ---RGTLR.N--.QW.-I--.....VLGFW-----NGG..GN----------K---L--------ER----------I--G----------D-----V----------  119
C.SO.SO145A              ---RGILR.NW-.QW.-I--.....ILGFW------GG..GN----------R--K--------S---ER----------S---ML-G----------D----------  119
C.UG.UG268A2             ---MGIQR.NC-.QW.-I--.....ILGFWI----NVM..GN----------E----------I--E----S---------D---Q-V----------  119
CONSENSUS D              -rvre??r.Ny-.hl.-k--.....i----m--tcs?a??enl---------v----------s-et-a-nI----------q-Iele----------N--e-----i------s-----  116
D.CD.84ZR085             ----GIKR.N---.P-.-K--.....I-------MTYSAA.DN----------S-K--A--I----------IE-K----------N--D----------  120
D.CD.ELI                 --ARGIER.NC-.NW.-K--.....I----I--T---A..DN----------S-E--A--I----------IA-E----------N----------  119
D.CD.JY1                 ---MGIRM.N---.--.-K--.....I----I--T--VA..-D----------S-EP-A--I----------R-IEME----------N----------N-----  119
D.CD.NDK                 --AR-KER.NC-.N-.-K--.....I-------T---A..-D---------I----------KK-A--I----------IE-E----------N----------  119
D.CI.CI13                -KAR-MNR.N--.--.---.....I-------T--VA..-N----------S-E--K--I----------IK-E----------N----------  119
D.SN.SE365A2             --AR-MKR.N--.--.---.....-------T--VA..----------S-E--K--I----------IE-A----------N----------  119
D.TZ.87TZ4622            ---R-MRR.N---.--.---.....----------..-Q----------S-E--A--I----------IK-E----------N----------  119
D.UG.92UG024-D           ---R-MQR.N--.--.-K--.....I----------G-.E-QS----------$KA-A--I----------R-IQ-E----------N--------V----------  119
D.UG.WHO15-474           ---RGIQR.N--.--.-G--.....-------T--TA..--K----------S-K--A-DI----------R-IE-E----------N----------  119
F1.BE.VI850              ---RGMQR.NW-.--.GK--.....LLF--I-I--N-A..DN----------ER-A----------F-K------D----N------T----------  119
F1.BR.93BR020-1          ---RGMQR.NW-.--.GK--.....LLF--T-I--N-A..-N----------S-EK-A----------E----R----E-N------T----------  119
F1.BR.BZ126              ---RGMQR.NW-.--.GK--.....LLF--I-I--N-...-N----------S-EK----------E------D----N------T----------  118
F1.FI.FIN9363            ---RGMQR.NW-.--.GK--.....LLF----I--K-A..DD----I----------N----------S-EK----------A-.---------E--------K----------  118
F1.FR.MP411              ---RVMQR.NW-.--.GK--.....LLF--I-I----A..DN----------G-ER----------IW-K------D----N----------  119
F2.CM.CA20               ---R-MQR.NW-.--.G---.....LLF--I-I----A..D----------ER--------Y-------S------G----------N--D----------E------  119
F2.CM.HIM277819          ---MGIER.N--.--.-K--.....-------LMTYSAA.DN----------S----------I-------Y-------S---LF-E----------N------A----------  120
F2.CM.MP255              ---R-MQR.NW-.--.GK--.....LLF--I-I--N-NATDD----------P----------P--------Y----------L--G---------E-N--D---L----------  121
F2.CM.MP257              ---R-MQR.NW-.--.G---.....LLF--I-I----A..D----------ER--------Y-------S---L--G--S-K------N--D----------E------  119
```

```
                                                signal peptide \/ gp120
                                                      <- vpu cds end
                                                      *                                *                               *                ^^^                                                                *
B.FR.HXB2                MRVKE....KYQ.HL.WRWG..WRWGTMLLGMLMICSAT..EKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCV 120
CONSENSUS G              -rvkGIqR.NW-.h-.-k-g.....tLi--LVI--sas..nn----------v-ED-dtt----s---aySt-s--v----------n---I?le-vt-n--m---n-----h-------dE------ 118
G.BE.DRCBL               ----GIQR.NW-.--.-N-.....ILI--LVI----...----------ED-NAP---------HS--S--I----------S---INMR----------N----------E------ 118
G.FI.HH8793-12-1         ----GIER.NW-.--.-K-.....-LI--LVI----S..NN----------ED-K----------S--R--I----------D---IP-G-------V----------E------ 119
G.GA.LBV217              --A-GILR.NW-.--.-K-.....ILI--LVI--N-S..GN----------A-ED-D----------S--S----------IS-N----------N----------E------ 119
G.NG.92NG083             ----GIQR.NW-.--.-K-.....-LI--LVI----S..DN----------ED-D-P--------S-SS-K----------IAIE----------N-----Q-------EE------ 119
G.NG.NG1937              ----GIQR.NW-.Y-.-T-W.....-LI--LVI---TS..NN----------ED-D------P-----S--R----------IT-E-I--I------N----------E------ 119
G.NG.NG1939              -K--GIQR.NW-.--.-T-W.....-LT--LVI----S..NN----------ED-D-P----------S--K----------IE-K--K--------N----------E------ 119
G.SE.SE6165              ---TGIQR.NW$.--.-K-.....-LI--LVI----S..NN----------ED-D----------S-SA-S----------IIME----Y------N----------E------ 118
H.BE.VI991               T--M-TQR.N-P.S-.----.....-LI----L---VV..GN----------K----------R----------M--E----T----V--------T----------      119
H.BE.VI997               T--M...R.N-P.QW.--G-.....IL-----L-Y--A..GN----------K----------EP-K----------S---M--A---------D-------QT----------      116
H.CF.90CF056             T--M-TQR.N-P.S-.----.....-LI----L----A..QN----------K----------E--K----------M-ME----S----E-N------T----------      119
J.SE.SE7022              T--M-TQT.SWL.S-.----.....L-IF--------R..-N----------RD-K----------S--K----------MS-P----------D--Q-----V--E------ 119
J.SE.SE7887              T--M-TQK.NW-.T-.--G-.....L-IF------K-K..-D----------D-K----------S--K----------S---MN-P----------D--Q-----V--E------ 119
K.CD.EQTB11C             --AR-IQR.NW-.---.GKR-.....ILF--I-I----A..NN----------E----------E----------N------T-------E------ 119
K.CM.MP535               ---RGMQR.NW-.T-.GN--.....ILF--I-I---NA..D----------P----------EK----------EME----------N------T-------E------ 119
N.CM.YBF106              ...MGMQS.GWP.....FFCL-ISLTI...GS..DPH----------RD-E-V---------HS--A--I---Q----------L-T----Y----E-K-A---Q-------E------ 109
N.CM.YBF30               -K-MGMQS.GWM.GMKSG-.....LLFY--VS-IKVIGS..-QH----------R--E----------HS--A--I---Q----------L-P----K----E-K-AD--Q-------E------ 121
CONSENSUS O              ????MkvMekkNkKl.w.....ilyiaMALiiP-Lss?.?qlYA---a-----ed-tPV---a--aNLTS--k--I--SQ-------t-?-yP-h---Dn-nI-k-Y--e--q-------e--l---v 114
O.CM.ANT70               ....MKAME-RNKK-.-.....TLYLAMALITP-LSL..RQ-YA---A-----ED--PV-------NLTS--K--I--SQ-------T-Y-YP-H---DD--I---Y-----Q----------      116
O.CM.CM4974              ....MKVME-KNKKS.-.....ILCIAMAL-IP-LNCX.NQ-YA---S------D--PV-------NLTS--K--I--SQ-------S-H-YP-E---DK-DI---Y--D----------      117
O.CM.HIV1CA9EN           ....MKVME-KNRK-.-.....ILYIVMALIIP-LSS..KQTYA---A-----GD-APV---T--VNLTS--K--I--SQ-------T-H-YP-H---DD--I-N-Y-----Q----------S----- 116
O.CM.MVP5180             ....MKVMK-NNRKS.-.....SLYIAMAL-IP-LSYS.KQ-YA---S-----E--APV-------NLTS--Q--I--SQ---------H-FP-G---D--DI---Y--D----------E------E 117
O.GA.VI686               ....MRVME-KNKK-.-.....ILC-IMALIIP-LSS..NH-YA---A-----ED-NPV-------NLTS--Q--I--SQ-------T-L-YP-K---D---I---Y-----Q-------E------ 116
O.GQ.193HA               ....MKAMERKNKKS.-.....IWYL-MALITP-LSS..KP-YA---A-----ED--PV-------NLTS--K--I--SQ-------T-Y-YP-H---DK--I---Y-----Q-------E--F---- 116
O.SN.MP1300              -T-TMKVME-RNRK-.G.....ILC-VMALITP-LSH..NQHYA---A-----E---PV------VNLTS--Q--I--SQ-------S-Y-YP-K---D---I-E-Y-----Q-------E------- 120
CPZ.US.CPZUS             -K-....ME-K.KR-.-.....LSYCL-SSLIIP.GLS...S--A---------RDVE----------KQ-A--I---Q----------H-P----K-D--E-N-A---Q-------------I 114
CONSENSUS 01             -rvk?Tqm.NWP.n-.-K-g.....-LIl-LVIi---S..dn-----y------rD-d-t-------k-he-----v----a----------IhlE------n----n-v---Q--V-----q--k--- 118
01_AE.CF.90CF11697       -G--GTQM.NWP.--.-K-.....-LI--LVI----S..DT----------RD-D---------N-QE----------V----------IH-E------D----N-----Q--V---------- 119
01_AE.CF.90CF402         ----GTRR.NWP.N-.-K-.....-LI--LVI----S..DN----------RD-D-I---------HV----------IY-E------D----N-----Q--V---------Q--- 119
01_AE.CF.90CF4071        ----GTQM.NWP.N-.-K-.....-LI--LVIM---S..DN------$-----RD-D----------HE-----I----------ID-E----------N-----Q--V---$---------- 117
01_AE.CM.CA10            --ARGTQM.NWP.N-.-K-E.....-LIF-LVI----S..D------H-------D-D----------HE----------IH-E----------N-----Q--V-----E------ 119
01_AE.TH.93TH065         -----TQM.N$P.N-.-K-.....-LI--LVI----S..DD----------RD-D----------HE----------IH-E----------N-----Q--V---------- 118
01_AE.TH.CM240           -----TQM.NWP.N-.-K-.....-LI--LVI----S..DN----------RD-D----------HE----------IH-E----------N-----Q--V---------- 119
01_AE.TH.KH03            -----TQM.NWP.N-.-K-.....-LI--LVI----S..DN----------RD-E----------HE----------IH-E----------N-----Q--V---------- 119
01_AE.TH.TH022           -----TQM.NWP.N-.-K-.....-LI--LVI----S..-N----------RD-D----------HE----------IHME----------K-A---Q--V---------- 119
CONSENSUS 02             ---MGiq?.nyp.ll.-r-.....?tiifWimi--n-...e?----------wrd-e----------dt----------n-q-ih-k----k------n------e---------- 116
02_AG.FR.DJ263           ---MGMQR.N-P.L-.-K-.....-IIFWIMI--N-...----------RN-E----------A----------IH-K----K------N---------- 118
02_AG.FR.DJ264           ---MGIQR.N-P.LF.-K-.....MIIFWIMI--N-...----------RD-E----------V----------IH-K----K------N---------- 118
02_AG.NG.IBNG            ---MGIQK.N-P.L-.----.....-NIFWIMI--N-...-Q----------T-E----------IH-E----K------N---------- 118
02_AG.NG.NG1921          ---MGTQR.NCS.L-.----.....-II-WLII--N-...KD----------RD-D----------EK----------D-H-ML-K----------N---------- 118
02_AG.SE.SE7812          ---MGIQK.N-P.L-.----.....MIIFWIMT------...GN----------RD-E----------IH-G----D------S---------- 118
02_AG.SN.MP1211          ---MGILK.SCP.PF.----.....MI---WI-I--N-...-N----------RRD-E----------E----------S---MH-E----------N------V---------- 119
03_AB.RU.KAL153-2        -----IRK.....--.----.....-LF----------..-N----------SK--------Y-------S---IP-K--------G--N---------- 116
03_AB.RU.KAL68-1         -----IRK.....--.----.....-LF----------..-N----------SK--------Y-------S-$-IP-K--------G--N---------- 115
03_AB.RU.RU98001         -----IRK.....--.----.....-L----------..-N----------SK--------Y-------S---IP-E--------G--N---------- 116
04_cpx.CY.94CY032-3      --MGMQR.N-P.--.-E-.....-LI--LVI----S..NN----------RD-E-------E----EK----I----------A-I----------NEG----A 119
04_cpx.GR.97PVCH         ---MGIQR.N-P.--.-E-.....-LI--LVIM---S..KDM----------RD-E-KPX---------K-I--I----------LA-T---------E-ST----------EG----- 119
04_cpx.GR.97PVMY         -T-MGTQR.NCP.R-.-T-.....-FI-WLVI----S..NN----------RD-E-SP----------K----I----------IF-K----D------N----------EEG----- 119
05_DF.BE.VI1310          ---RGMQR.NWP.--.GK-.....LLF--I-I------..D-F----------R--K----------G--K----------G----------T---------- 119
05_DF.BE.VI961           ---RGMQR.NW-.--.GK-.....LLF--I-I----A..DN----------N----------G-EK-A----------LA-E----------N------A----------S-- 119
06_cpx.AU.BFP90          ----GIQT.SW-.--.-K-.....-LI--LVI----S..KNM---------A-ED-D-I----------SPDK----------IS-K----------N--D----------E------ 119
06_cpx.ML.95ML127        ---RGIQK.NW-.--.-K-.....-LI--LVI----S..NN----------A-ED-D-I----------SA-K----------IP-K---------E-H----------E------ 119
06_cpx.ML.95ML84         ----GIQM.NW-.--.-K-.....-LI--LAI------..SN----------ED-D-I----------SA-K----------IP-E----------H--Q----------E------ 119
06_cpx.NG.NG3670a        -T--GIQK.NW-.--.-K-.....-LI--LVI----S..NNM---------A-E--D-I----------S--K----------IT-E----------H----------E------ 119
06_cpx.SN.97SE1078       -T--GIQR.NW-.--.-K-.....-LI--LVI----S..QNM---------A-ED-D-I----------SA-K----------IK-E----------N----------EE------ 119
10_CD.BFL061             -T-TGIQR.NCR.QW.-I-.....ILGFW-----N-R..----------T----------KA-A--I----------R-I--E----------G--D----------G----I 119
10_CD.BFL071             -K-MGIQR.N-P.QW.-I-.....ILG-W---.ICN...GN----------Y-KE-A--I--A----------L--E-I----------D---------- 117
10_CD.BFL110             ---MGIQK.NC-.PW.-I-.....ILGFW-----NV-..G----------T----------G-K-----I----------I--E----------G--D----------G----- 119
CONSENSUS 11_cpx         -rvk?tQr.NWh.n-.-r-.....L-If------N-t..dnl----------rD-?-t-------kaySt-K------ha------n-q-ip-e----n--m---n--e--h--i-----E-----v 117
11_cpx.CM.MP818          --ARGTQK.NWH.D-.----.....L-IS------N-...DN----------D-N---------Q--SP-K-------Y----------IL-G----------N----------E------ 119
11_cpx.CM.CA1            -K-RGSQM.NWH.N-.-K-.....L-IF------N---.D----------RD-D----------S--K----------D---IY-E----------N----------E------ 119
11_cpx.FR.MP1298         -----TQR.NWH.N-.----.....L-IF------N---..DN----------RD-N----------VS--K-------S----------IF-E----------N----------E------ 119
11_cpx.FR.MP1307         -----TQM.NW-.N-.----.....L-IF------N-A..GN----------ED-N----------S--K------Y----------MP-K----E------N----------E------ 119
11_cpx.GR.GR17           -----TQR.NWH.N-.----.....L-IF------N-...--M----------RD-D-I-------RT-S--K-------S--------R-LS-E----------N--------V-----E------ 118
CPZ.CM.CAM3              .....ME-I.KGS.....LIEKY-FM-LIIP-LTGS.DQ--A----------D-------E-S-LNK-A--I--SQ----------QIP---------N-T-A---NT---------- 114
CPZ.CD.CPZANT            --.....KPIH..II.-.....GLALLIQFIE....KG-..NEDY---F------RN--P-----TN-SMTS---------TS---I--D-IV-R-..-TSVW--AY--Y---S-T--MXQ-FQ--H---- 110
CPZ.CM.CAM5              .....MERK.KLN.-.....IICY--M-LITP-LTG..HEW-A----------E-------E-S-LNK-A--I--SQ----V-------YMP--S-S------N-A---Q---S---------- 113
CPZ.GA.CPZGAB            -K-....ME-K.KRD.-..NSLSIITIITIIL-TP-LTS...E----------HD-DPV---------HS--A--I---Q-------S----F-P--I-S------N--D---------- 119
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                  V1                                  V2
B.FR.HXB2                KLTPLCVSLKCTDLKNDTN.........TNSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDN..............DTTSYKLTSCNTSVI.TQACPKVSFEPIPIHYCAPAGFAI     225
CONSENSUS G              k-----vT-N--nvnn???..?????t????t?nn?tvt?eee?k----niT-Eirdkkkk---l--rl-Vvpind......??????n?n?t?-R-in--v-t-.K------t-d---------a----     214
G.BE.DRCBL               -------T-N--EIN-NST.................-N-T-EYRMT-----MT-EL-D-KKA---L--RT-VV--NE......MNNENNGTNS-W-R--N--V-T-.K------T----------------     225
G.FI.HH8793-12-1         -------T-N---ANVT-V.................ANESVSAQ---------T-E--DRKR----L--R--V---ND.......DSSNSTGNYSN-R-IN--V-T-.K------D-D---------G----     226
G.GA.LBV217              -------T-N--NAIANVT...NSNKTSNNS-LGNSTVSSIE-M-------T-EV-D-SK----L--R--VV--NG..........NENSS--D-R-IN--A-T-.K------T-D----------------     235
G.NG.92NG083             ------IT-N--NVNSANH.............TEANNTVEN-E-------K-T-ERG--KKE---L-----VV--S-..............GNK---R-IH--V-T-.K------N-D----------------     223
G.NG.NG1937              -------T-N--SDN-VNC.......TNANCTNNSTRSL-EANLR------T-E--DRQK----L---P-VV---D.......QSNSSNSSYS--R-IN--V-T-.K--------D----------------     234
G.NG.NG1939              R------T-N--NVNTTSC.......TNN-TTKNTTVDTMCDSMT-----VT-E-KDRQK----L--R--VVQFKD.............NDNSSK-R-IN--V-A-.K--------D----------------     229
G.SE.SE6165              -------T-N---VT-KG-..KRNNSTDNSTETNNSTVDNP---------VT-E--D-KK-------R--VV--N-..................A-N-R--H--V-T-.K------T-D---------T----     230
H.BE.VI991               -------T-D-SSVNATNV........TKS-N-TDIN-G-IQ-QR-----VT-A--D-N--VH-L--RA--VQ--E.........GERNKSDNH-R-IN-------.K----------------------     231
H.BE.VI997               -------T-D-SNITRNDT........NSSSTVNATSSPSAN-LT-----VT-V--D-Q-RVH-L--R--VV---E.......TSNNNNSNS-K-R-IN-------.---------D----------     230
H.CF.90CF056             -------T-N--NVR-N-S.............N-TSSMEAG--LT-----VT-VL-D-Q--VH-L--R--VV----.............NS-Q-R-IN-------.----------------------     221
J.SE.SE7022              -I-----T-N-S-VNSNNS...........-D-N-SASNNSPEIM------VT-E--N-RKQ---L--RQ-VV--NS..............-NK--I-IN-------.K--------Q----------     224
J.SE.SE7887              -I-----T-N-SNITSNS-...........TT-NSSVSSPDIMT------T-E--N-RKQ---L--RQ-VV---S..............NNKN-I-IN-------.K--------Q----------     222
K.CD.EQTB11C             -------T-T--NVT-NRT.....NANKNDTNINATVTSTD---------T-ELKD-KKRVS-L------VQ-KQ.........SEINQSESEDR-IN----TV.----------------------     233
K.CM.MP535               E------T-N---Y-GTNS...........TNNATSTVVSPA---------T-E-KD-KK--S-L--R--VL-LNG.........EGNNSS-E-R-IN----T-..--T----T----------------     227
N.CM.YBF106              -------TML-NNSNGNSA.........GNSTTNRTEDL-DRQM-------T-E--DRKKQV-SL--VE-VV--KD..........GT-NNT-R-IN---TAV.------TT----------P----     217
N.CM.YBF30               -------TML-N-SYGEER........NN--MTTREPDIGYKQM------AT-ELTD-KKQV-SL--VE-VV--NA..........YNKT-R-IN---TAV.------T----------P----     228
CONSENSUS O              qM-F---QMn-tn??gtt?....??????????????s?enlmkq-e--vT-VlkD-KEkkQ-L--vs-L?klnn.......????n??nstm-t-iN--STt-?k-a-------p---------a-y--     206
O.CM.ANT70               QM-F---QME--NIAGT--.................ENLM-K-E--VT-V-KD-KE-KQ-L--VS-LMELNE.......TSSTNKTNSKM-T--N--STT-.----------------------Y--     217
O.CM.CM4974              QM-F---QMN--NAQLRII.................QYK-HXM-R----TT-VFKD-KEQKQ-L--VS-LTA-HD.........KNNNK-I-T-IN--STT-XK----------------------     218
O.CM.HIV1CA9EN           QM-F---QMN--TPPQPNS..........-S--ENVTSIPENPM-Q-E---T-VLKD-KE-KQ-L--VS-LVKLN-..........TGNGND-M-T-IN--STT-.K----------------------Y--     225
O.CM.MVP5180             -M-F---QMN-V--QTNKT.................-LLNETIN-MR-----VT-VLTD-KEQKQ-L--VS-LSKVND.........SNAVNG-T-M--N--STI-.K------------------T-Y--     221
O.GA.VI686               QM-F---QMN--NIS-A-S....SNISNA-S-DISKDTSSENSV-Q-E---T-VLKD-KE-KQ-L--VS-LVKT--..............S-M-T-IN--STT-.K----------------------Y--     226
O.GQ.193HA               QM-F---QMN--NPNGTAA.................NSTENHL-Q-E--VT-VVKD-KERKQ-L--RA-L-KLE-..............S-M-T-IN--STT-.K----------------------Y--     212
O.SN.MP1300              QM-F---QMN--NVNDE--...........S-VKNDTSSSENLM-K-E--VT-VLKD-KE-QQ-L--VS-LM.............KVNENND-M-T-IN--STT-.K-T--------A----------Y--     225
CPZ.US.CPZUS             -------TMT-LNPDSNSS.................AVNTTDIMR------T-EL-D-KKQV-SL--VD-LAH-N-..............NT-R-IN---TA-.-------T----------P----     210
CONSENSUS 01             k------T-n--?anltn?....??????tnnv?kiignitdeVr--s-nmT-ElrD-k-?vh-l--kl-iVqi?g?????????????nnsseyr-In--t----.K-----vs-d-------t---y--     219
01_AE.CF.90CF11697       -------T-H--KA-LNDT..........Y-GTAKLNDTIGD-VR-----VT-EL-D-K-EVH-L--VP--VR-GE.......KNKNSSGNSSE-I-IH-------.K------T-D-------T---Y--     231
01_AE.CF.90CF402         -------T-H--KASFTNA.............TSDR-KMEDAVR-----MT-ELQD-K-EVH-L--TS-VVQ-SSSVQNNNNSNTSGQNNSHKFR-IH-------.K--------D----------Y--     234
01_AE.CF.90CF4071        -------T-D--KADFY-T.....KFNT-EKPEIEISEMQK-VS--P---T-EL-D-E-EV--L--RS-LV--ER.........NSGENNGS--R-IH-------.K--------D----------Y--     231
01_AE.CM.CA10            -------T-N--KANFNNT....STGNSTDEDKFKSIGNITESV-----DMT-E--D-K--VH-I------VQMGG.......QSSNNSNGYNK-R-IN--S----.K----------------------     237
01_AE.TH.93TH065         -------T-N--NANLTNV.........N-NV-NIIGNITD-VR-----MT-EL-D-K-QVH-L------VQMNK.............NSSE-R-IN-------.K-----I--D-------T---Y--     223
01_AE.TH.CM240           -------T-N--NANLTNG........SSKTNV-NIIGNITD-VR--T--MT-ELTD-K--VH-L------VQ-ED............KKTSSE-R-IN-------.K-----I--D-------T---Y--     228
01_AE.TH.KH03            T------T-H--NANLTDV.......TE-TNVPKIIGNVTD-VR----KMT-EL-D-Q-QD--L------V--EH............KNNSSE-R-IT-------.K-----I--D-------T---Y--     228
01_AE.TH.TH022           -------T-N--NVNATNV.........S-TTEAPN-VGTD-V------VT-EL-D-T-QVQ-L------VQMGG.............N-SGE-R-IN-------.K--------D-------T---Y--     225
CONSENSUS 02             k------T-d-hnvnss?s.??????????????nissdmqg-i------mT-elrD-Kqkvs-L--rl-vvq-ne....?????????s?nsq-R-IN-n--a-.tq---kvsf------------f--     215
02_AG.FR.DJ263           -------T-D-HNVNSSNS.STSNSSNSS-PINRTIDSDMQE--------MT-EL-D-K--VS-L--R--VVQ-NE.............SGNSQ-R-IN-H--A-..K---R-TL-----------L--     233
02_AG.FR.DJ264           E------T-D-YNVSSNSS...........I-VKSISRDMQ---------MT-EL-D-K-QVS-L--R--VVQ-NE.............SSNSQ-R-IN----A-.K------T----------------     223
02_AG.NG.IBNG            -------T-D-HNFN-SYS............N--NLTSDMN----------T-EV-D-KK-MH-L--R--VVQ-NE.............NNGSQ-R-IN----A-.----------------------     222
02_AG.NG.NG1921          -------T-D-YNVNSSYS.......NSSSHNNRSISSEMQR--------MT-LL-D-KE-VS-L--RQ---K-NK....NENGSEETETSSNQ-R-IN----T-.------I----------------     236
02_AG.SE.SE7812          Q------T-H-Q-NLTSSG.................NISENMQ---------MT-EL-D-K--V--L--RY-VVQ-NE.............TG-NIQ-R-IN----A-.----------------------     220
02_AG.SN.MP1211          -------T-E-HNYNYTR-.............NSK-N-VQ-M------MT-VLKD-KK-MA-L---I--V---K..............NA-Y-R-IN----A-.----------------------     219
03_AB.RU.KAL153-2        -------T-N-----KEVT.............-TNTSSIKMM-M-------T-DL-D--K----L-----VVQ---.................D--R-I------V.------I----------------     216
03_AB.RU.KAL68-1         -------T-N-----KNVT.............-TNTSSIKMM-M-------T-DL-D--K----L-----VVQ---.................D--R-I------V.------IF---------------     215
03_AB.RU.RU98001         -------T-N--EV-TNDT.............STNASGIEMM-------T-DL-D--K--H-L-----VVQ---.................D--R-I------V.------I----------------     214
04_cpx.CY.94CY032-3      ---S---TFT-INATTTNS.............TNGTVI-EG------D-T-E--D-KK----L--RI--V--NA...RVPINGSNRNNS-EE-M-IN--A-T-.K----------------------     230
04_cpx.GR.97PVCH         -------A-N-SNATIII-....STNNST---TGNSTVKSTA---------T-EV-D-QK--H-L--R--VV--N-...NVPINNTSNTSEYRE-R-M-----NH.K------TL--NSHTLLC-GW-CD     241
04_cpx.GR.97PVMY         -------A-N-G-ATIKST.............N-TPNVTTTN-M-------T-E--D-KK-A--L--R---V--ND....NNSTNSRRSSN-SD-M-IN---V-S-.------IK----------------     231
05_DF.BE.VI1310          -------T-N---F-ANST........ANSTTN-TTLKE-T-AVQ-----MT-EVND-KL-VH-L--R---V--SS..........DDSSNS--R-IN----T-.K-------WD----------Y--     229
05_DF.BE.VI961           -------T-N---ATSA-T.TSKNISATP-SNPNDTLKE-Q-A-Q--T---T-EVKD-NKRVH-L--R---VS-NS.............SRKE-R-IN----T-.K------CWD----------Y--     233
06_cpx.AU.BFP90          -------T-T--NATLGNKTLGNNSTNSTLGNN-TIVDDISK---------T-E--D-TK----L--RP--V--GD............DSNNSD-R-IN--V-T-.K----------------------     236
06_cpx.ML.95ML127        -------T-I--NINITST.............--TKNITVESGE--------VT-E--D-QKE---L--R--VV--ND...........GSNNN--R-IN--A-T-.K------T----------------     225
06_cpx.ML.95ML84         -------T-N--NVTDHGI.............NNNNTVEG-E--------VT-E-KD-KK--R-L--T--VV--ND...........NGNNST-R-IN----T-.K----------------------     223
06_cpx.NG.NG3670a        -------T-Y--NVTKNNS...........TGNNSSTVEDRE-MT-----VT-E--D-QK----L-----VVQ--G.............SNKT-R-IN--V-T-.K------T-G----------G-     223
06_cpx.SN.97SE1078       -------T-N--NVNATKY........SNETVGKSLTVKDRE---------T-EV-DQKKT---L--RP-VMQV-G.............KNSST-R-IN--A-T-.K----------------------     227
10_CD.BFL061             -------T-N-NGINATND.................TVGDPRMT-----TT-EVKD-KKQ---L-----VVQ-GD.............SN---R-IN----A-.-------T----------------     216
10_CD.BFL071             -------T-N-S-VASNA-...........VT-GTNSTDSDTGM-------T-E--D-QK--H-L-----VVQ--G.............SN---R-IN----A-.K------T----------------     221
10_CD.BFL110             -------T-N-SEWNATNH...........EGRTVNATNNVGM-------T-E--D-KKQ---L--RS-VVQL-G.............NTN---R-IN-------.-------T----------------     223
CONSENSUS 11_cpx         -------t-n-tdv?snnt.......???????????v???a?mk-csfnvttel?dkkkr-yalf-kldvVp-nd.......????nsn??d-rlin--vstv?K-acp--t-------------f-i     208
11_cpx.CM.MP818          -------T-N-AEVTSYNT.....................T-M------VT-ELIDRRKQ---L------V--ND.........NNNSSNVSD-R-IN--V-T-.K------T----------------     217
11_cpx.CM.CA1            -------T-N---V-TKNR.................TVEDQA-L------VT-ERNN-KKR---L-----VV--NA.........NNNSTAGD-M-IN--V-TV.K------T----------Y--     221
11_cpx.FR.MP1298         -------T-N--N-NVTNS.......TAVGADVTNS--E-----E----RVT-EVTY-NEV---L---H-VV--NV..............N-T-R-IR--V-AV.K------T----------------     226
11_cpx.FR.MP1307         -------T---ANVTSNST.............ADTTEVDRQDDML-GTLS....LPERKLE-NTHE---FKVL-KV............NIKVND-NINK--IVT-NK-GS---T----------X     220
11_cpx.GR.GR17           -------A-N---ARDNAT.....................D----T---T-ELED-KKN-R-L--R--VV--ND.........SSSNIGQ-R-IN--V-AV.K---S--T----------------     213
CPZ.CM.CAM3              ----I--TME-RKVTFNST...SNRNKTS-MTTNSPNEKXDSTV---T--MT-EV-D-EK-V-SL--VD-LVL---.............--DT-R-IN---TA-.------T----------TP-Y--     226
CPZ.CD.CPZANT            ----M-IKMN--GYNGTPT..TPSTTTSTVTPKTTTP-VDGMKLQE-N--Q--GFKD-K--MK-I---G-LMKCQD............NNE-NC-Y-WH---TT-.--S-E-ST----------Y--     225
CPZ.CM.CAM5              ----I--TMQ--T---IST...........--TN-TELKNI-EIM---T--MT-EV-D-KKEV-SL--VD--V--GG..........NKSNDT-R-IN---IVT-.------T----------P----     220
CPZ.GA.CPZGAB            -------T-Q-SKANFSQA...........KNLTNQTSSPPL-M------VT-EL-D-KKQV-SL--VE-VVNLG-..........ENNT-RIIN---TA-.------T----------------     223
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

V3 neutralizing loop

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins


```
                                CD4    |-      V5^^^     -|                                                
```

```
                    CD4   |-      V5      -|                                                gp120 \/ gp41
                          |                                                                  |-  fusion peptide  -|
B.FR.HXB2           TRDGGNS......NNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVV.....QREKRAV.G.IGAL.FLG.FLGAAGSTMGAAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLT 569
CONSENSUS G         -----nn..?????st?-t---g--d-rd---s------?vk-Kp--v---r-R-------.....e-----v.g.l--v.fl-.---a---------it--v-V----------s-------a-------- 533
G.BE.DRCBL          -----DN.......-ST-----------KN----------T---KS--I---R-R-------.....E-------.-.V--I.---.---T---------I-----V----------S---------------- 560
G.FI.HH8793-12-1    ------A....NNAS-T-T---A--N-K----N----------K---------R-------.....G-------.-.V--V.---.------------I-----V----------S---------------- 556
G.GA.LBV217         ----------.....T-GN-T---A----------------I--K---------R-------.....G-----I...T--V.---.------------LA----V---------------------------- 570
G.NG.92NG083        ------N.....NDSTE-T--------------------T---KS------R-R-------.....E-------.-.L--V.---.------------I---A-V----------S---------------- 555
G.NG.NG1937         -----TN.......KTN-T---A----------------R-K---I-----R-------.....E-------.-.---V.LF-.------------I-----V----------S-------V-------- 576
G.NG.NG1939         ------I........TE-T--------------------I---K-------R-R-------.....E-------.-.L--V.L--.---T---------I-----V----------S---------------- 559
G.SE.SE6165         ------N..NNTNTS----------------------T---KS------R-R-------.....E-------.-.L--V.---.------------I-----V----------G---------------- 572
H.BE.VI991          -F-RN-T........NNVT-------------------------------E-R-------.....E-------.-.M--F.---.------------I----------S------Q----M---- 575
H.BE.VI997          -Y-EDKG.......--DNVT-----------------------------E-R-------.....E-------.-.M--F.---.------------I----------S------Q----M---- 568
H.CF.90CF056        -I-E--A......SA-NYT-------------------------I----TR-------.....E-------.-.M--S.---.------------I----------S------Q-R--M---- 561
J.SE.SE7022         ------R...NGSE-GT-T---T--N-K------------EL----------------.....E-------.-.---V.---.---T---------I-----V----------S---K--X------K-- 566
J.SE.SE7887         ------R..GNGSE-GT-T---T--N-K------------E-----------------.....E-------.-.---V.---.---T---------I-----V----------S---K---------K-- 564
K.CD.EQTB11C        ------D....NNTRTE-T---------------------Q-----I---R-R-------.....--------.-.---.---.------------I----------------------QM---- 564
K.CM.MP535          ------N.......THN-T---------------------Q-----I---R-R-------.....--------.-.---.---.------------I----------S---------------- 555
N.CM.YBF106         EHS-E.........-GNKTVY-S--N-VNL--Q--------S---I----G-----T-.....S-----AF-.L---.---.------------I-------T----------------------S 540
N.CM.YBF30          EYS--PD.......TKET-VY-S--N-VNL--Q--------S---I----G-----T-.....S-----AF-.L---.---.------------I-------T---------I----------S 557
CONSENSUS O         qldnpwN....??sn???tfr-i----k-I--te-fn----rvK-FS----kia-PvIgtgt?h-------.-.l-Ml?---.V-S---------ATt-a-qtHtlmK------Dn-----Qa--q--R-S 536
O.CM.ANT70          QM-NTWN.....SS-NNVT---I----K-I--T--FN----RVK-FS----RIA-P-ISTRT.H-------.-.L-M-.----.V-S---------AT--A--THT--K------D------Q---Q--R-S 567
O.CM.CM4974         ELHNPYN.....KS-PNTT-G-T------I--A--CN----QVK-FS-----VS-PIISLNNPH-------.-.L-M-.----.V-S---------AT--A-RTHS-IK------DS-----QT-----R-S 576
O.CM.HIV1CA9EN      QL-RPWN......SSRDNTL--A------I--T--F------VK-FS-----IA-PTIGTGTPH-------.-.L-M-.----.V-S---------AT--A--THT-MK------D------Q---E--R-S 578
O.CM.MVP5180        QL-QPWN......STGENTL--V----K-I--TK--N----Q-K-FS-----MS-PIINIHTPH-------.-.L-M-.----.V-S---------ATA---RTHSV-K------D------Q------R-S 577
O.GA.VI686          QM-MPWN.......KSNTT---V----K-I--T--F-----RVK-FS-----IA-P-IGTGT.H-------.-.L-M-.----.V-S---------ATA-----HT-IK------D------Q---Q--R-S 578
O.GQ.193HA          QL-NQWN.....TSTNEAK---I------I--K--FN----RVK-FS----HIS-P-IGTRT.H-------.-.--M-.----.V-S---------ATA----THT-MK------D------Q---Q--R-S 564
O.SN.MP1300         QL-KTWN....RT--SEST---I----K-I--T--F-------K-FS-----IA-P-IGTGT.R-------.-.L-M.L---.V-S---------AT--A--THT-MK------D------Q---Q--R-S 575
CPZ.US.CPZUS        DISISAV....NNDSRNITVM-T----TAL-KN--H-----S---I----G----HT-.....K-----AF-.L---.---.------------VV----------------------------S 541
CONSENSUS 01        t----?n...???nttn-T---g--NiK----s-------vq-----i--tr-k-------.....E-------.-.---m.IF-.------------i---------l-------S----a---------- 548
01_AE.CF.90CF11697  ------A........TN-T------N-K------------Q-----I---R-R-------.....E-------.-.---M.IF-.------------L----------------S---------------- 569
01_AE.CF.90CF402    -----IN...QNQT-KN-T------NIK------------Q-----I-----R-------.....E-------.-.---M.IF-.------------I----------------S---------------- 577
01_AE.CF.90CF4071   V----KI.....E-ETI-T------N-K------------Q---------G-------.....E-------.-.---M.IF-.------------I----------S----T---------- 564
01_AE.CM.CA10       -----GN......STDT-T------N-K-----------IQ-----I-----R-------.....E-------.-.-----.IF-.------------L---------M-------S---------- 575
01_AE.TH.93TH065    -----SA.....N-TNN-T---E--NIK------------Q-----I---R-------.....E-------.-.---M.IF-.------------I----------------S---------------- 560
01_AE.TH.CM240      -----VN.......-TDN-T------NIK------------Q-----I---R-------.....E-------.-.---M.IF-.------------I----------------S---------------- 563
01_AE.TH.KH03       -----PN.......-MTN-T------NIK----N--------R-----I---R-------.....E-------.-.---M.IF-.------------I----------------S---------------- 564
01_AE.TH.TH022      -----DN.....N-TIN-T------NIK------------Q-----I--S----------.....E-------.-.---M.IF-.------------I----------------S---------------- 563
CONSENSUS 02        -----?n......nstn-T-------------------------p--v--tha-------.....E-----v?-.L--v.-l-.----------a-I----------------s---?---------k-- 549
02_AG.FR.DJ263      -----SN.......-STN-T-------------------------------RP-------.....E-------.-.L--V.---.------------I----------------S----------K-- 575
02_AG.FR.DJ264      -----SN.......-STN-T----------------------------I---H-------.....E-------.-.L--V.-F-.------------I----------------S----------K-- 561
02_AG.NG.IBNG       ------N.......STN-T-----------------------------------R-------.....E-------.-.L--V.---.----------R-I----------------S---K---------K-- 560
02_AG.NG.NG1921     -----KH.......-STN-T--------------------------A-------Y-------.....E-------.-.L--V.---.------------I----------------S----------K-- 579
02_AG.SE.SE7812     ------D........--NT-T-------------------------------SH-------.....E-----LV-.L--F.-F-.------------I----------------S---K---------R-- 572
02_AG.SN.MP1211     -----DN......T-GD-T-------------------------------SH-------.....E-------.-.L--V.---.------------I----------------------K---------K-- 566
03_AB.RU.KAL153-2   ------Q......S-VT----T-----------------------------R-------.....--------.-.---V.---.------------I---------------------------------- 553
03_AB.RU.KAL68-1    -----YQ......S-VT-------------------------------------R-------.....--------.-.---V.---.------------I---------------------------------- 552
03_AB.RU.RU98001    ------Q........--VT------------------------------------------.....--------.-.---V.---.------------I---------------------------------- 551
04_cpx.CY.94CY032-3 -----TN........NT--------------------------I----N--R-------.....----$--.-.---M.---.---------------------------------S---------------R-- 567
04_cpx.GR.97PVCH    -----I-..NNNET-DN-TL-------------V-----------V-I---G-R-------.....-K------.-.L---.---.------------L----------------S---------------K-- 588
04_cpx.GR.97PVMY    ------R.....SDTGN-T-----------------------Q---V------R-R-------.....--------.-.---V.-P-L----------I----------F------S----------Q--R-- 574
05_DF.BE.VI1310     A-----G...NDSS-DT-----------------------E-Q---I---R---Q--.....K-----M.-.---M.------A----------IA----------------------Q------------ 574
05_DF.BE.VI961      -----VN.....IT-DT-T----------------------E---------R---Q--.....--------.-.V---.LI-.---------------------------------Q------------ 566
06_cpx.AU.BFP90     ----N-E.......-V--T-----------------------K---I---W-R-------.....G-------.-.L--M.---.---T---------I-----V----------S---------------- 572
06_cpx.ML.95ML127   -----RN......DSN--T------------N--D-------K---I---R-R-------.....GK-----V-.L--V.---.---T---------I-----V----------S---------------- 558
06_cpx.ML.95ML84    ----NTN.......TT---I------------N----------K---I---E-R-------.....G-------.-.---F.---.---T---------I-----V----------S---------------- 553
06_cpx.NG.NG3670a   ----N-D.......SHT-T----------------------K---I---W-R-------.....G-------.-.L--V.---.---T---------I-----V--------H--S---------------- 553
06_cpx.SN.97SE1078  ----H-D.........T---------N-K------------K---I-----R-------.....G-------V-.L--V.---.---T---------I-----V--------H--S---------------- 559
10_CD.BFL061        -----AN.....N-SQN-T----------------------I-L---T----------.....E-----I-.-L--V.---.------------L--------V---------------------------- 553
10_CD.BFL071        ---RDGG...ANNTSQN-T--------------X-------------L----------.....E-----I-.-L--M.---.------------L----------------------K------------ 565
10_CD.BFL110        -----AN....NS-QSG-----------------------R-----L----------.....E-----I-.-L--V.---.------------L--------------------------M----- 561
CONSENSUS 11_cpx    -----nn....???ttn-T---T--d-rd---s---k----E-k-------r----------.....e-----v?-.i--v.L?-.---a---------I-----a----------sn--k--ea--?l-K-- 529
11_cpx.CM.MP818     -----LN.......STN-T---T----------------E-K-------R----------.....E-------.-.---V.LF-.------------I-----V----------S---K---------K-- 558
11_cpx.CM.CA1       ------K.......TQN-T---T----------------E-R----------.....--------.-.---V.LF-.------------I----------------S---K------Q--K-- 560
11_cpx.FR.MP1298    -----PN.......-TT--T---T--N----------R----E-K-------R----------.....E-----V-.L---.LF-.------------I-----T----------S-------------K-- 567
11_cpx.FR.MP1307    ------M.....NDTD--T---T--N------N--------E-K-------R----------.....E-----AV-.L---.LF-.------------I----------------------K------QM-K-- 563
11_cpx.GR.GR17      -----EG....NDTIGK-T---T-----N-----------E-K----------.....E-------.-.---V.L--.------------I----------------------K---V-----K-- 555
CPZ.CM.CAM3         ETGH-IN.....NSITNITLY-T--N-V-L--L--H-----S---I----S----HT-.....S-----AF-.L---.---.------------VV----------------------------S 561
CPZ.CD.CPZANT       EGQIY-E.........TVKVS-.AARVA-Q--A--SR-Q--E-X--S----TX--PEIKQH..S-Q--GI.-.---LF.---.L-S----------IA--A-T-N-XH------A---Q---T-------S 562
CPZ.CM.CAM5         EAG-KWE.......-NTITVY-S--E---L--L--NR----S---I---------HT-.....A-----AF-.L---.---.------------VV------H----------------------S 567
CPZ.GA.CPZGAB       -S-TPVT.....N-SGNLT---T--N-K-I----------R----S------R-HT-ARQK.D-Q---AF-.L---.---.------------AV----------------------K----------S 559
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins


```
                                                  immunodominant region
                                                  |-        *     * -|   ^^^   ^          ^^       ^^^          ^^^                                    ^^^
B.FR.HXB2          VWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWS.N.......KSLEQIW..NHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMI    688
CONSENSUS A        ----------Vl-v----r---l---------l--t-n----Ss--.-.......-sqse--..?NM--lQ-dk-isn--di-Yq-i-e--nq--k--qd--a-----n--n--d-sn-----ki-im-    663
A.GB.MA246         ----------V--L----R---I-----------N----S---.-.......--QSE--..ENM--LQ--K--S---HI-YT-L----I-------D--A-----N-----D---------I----    683
A.KE.K89           ----------V-------R-------------I----N----S---.-.......--QSE--..ENM--LQ--K--S---DI-YN-L-----------D--A-----N-----D-S-------I----    680
A.KE.Q23           ----------V-------R-------------------N----S---.-.......---DE---..-NM--LQ--K------Q--YR-----------K---------N--S--D-S-------I--I-    678
A.RW.PVPI          ----------V-----------------------P-T----S---.-.......-T-SE--..DNM--LQ--K--S---D--YQ---D----------D--A----------D-S------RI----    691
A.SE.UGSE8131      ----------VM--------------------I---------ST--.-.......--YT----..DNM--LQ--K--S---DI-YQ------Y------K---------N-----D-S-------I----    696
A.UA.ukr970063     ----------V-------------------------N----SN--.-.......--QSE--..DNM---Q----VIY--DI-YD---K---P------D--A--------S--D-S-------I--I-    686
A.UG.92UG037       ----------V-------R----------------P-N----S---.-.......---DE---..ENM--LQ--K--S---IK-YE------I---R--KD-------------D-SK------I----    682
A.UG.U455          ----------V-------Q----------------T----S---.-.......--Q-D---..-NM--LQ-EK--SS--GI-YQ-------------LD--A-----N-$-----S------R--VI-    680
A2.CD.97CDKS10     ----------V--L----Q----------------T----S---.-.......-TY-E---..-NM--LQ-----D---NI-YN-L-----------D--A-----------S--------RI----    686
A2.CD.97CDKTB48    ----------V--L----Q----------------A-D-R--S---.-.......-TQ-----..KNM--LQ--K--ST--DI-YM-L-----------D--A-----N-----D--R------I----    675
A2.CY.94CY017-41   ----------V----------------------A-T----T---.-.......--QDE---..DNM--LQ--K--S---NI-YR-L-----------D--A-----D--S----SH-----RI----    690
CONSENSUS B        ----------v-------k--q-l-i-----kl--t-a----a---.n.......-slde---..nNm---e-er-id--tsl-Yt-i-e--nq-----qe--e-dk-as--n-fd-tn-----ki--m-    675
B.AU.MBC18         ----------V-------R------L-----------T----T---.-.......-T--K---..DNM---Q-ET------NI-YT-L----------KD-----Q--N-----T-S-------I----    696
B.CN.RL42          ----------V-------------------------------------.-.......---HE---..-NM-----E---D---RE-YT-----------L---------------D--K------I----    677
B.DE.D31           ----------V-------R-------------------------------.-.......--MDM---..-NM-----E---D------YT-------------------N--EN--S--D-S-------I----    682
B.DE.HAN           ----------V-------R----------------------T--------.-.......-T-D---..-NM-----E---D------YT---Q---------------------S-YD-S-------I----    687
B.ES.89SP061       ----------V-------R------F---------------T---.-.......---S----..DNM---Q-E---D------YN--------------------N--------S---------I----    694
B.GA.OYI           ----------V-----------------------T--------.-.......---NE---..DNM---Q-E---D---H--YT----------------------G--S--S---------RI--I-    687
B.GB.CAM1          ----------V-------------------------------.-.......---DK---..-NM-----E---D---N--YT-------------KD-----T--------D---------I----    688
B.GB.WB            ----------V-------Q--R---------R--------------.-.......---NE---..DNM-----E---E---G--YN-L-T-----------------PN------------------    691
B.JP.ETR           ----------V-----------M--------F---------T----.-.......--FNE---..DNM-----E-------N--YN-----------D--A----D------S--K------I----    694
B.JP.JH32          ------------------------------------------------.-.......----E---..DNM-----E---D------YT---------------G----------T--------RI----    699
B.KR.WK            ----------V-------R----------------N----T---.-.......--ENE---..DNM-------------N--YD-L-K---------------------$------K------I----    697
B.NL.3202A21       ----------V-------------------------------------.-.......---DK---..-NM--------VS------YT-------------------------D--K------I----    688
B.TH.93TH067       ----------V-------------------------------------.-.......---NE-$..-NI-----E---D---RE-YT-------L-----L-----N---------D--K------I----    670
B.TT.QZ4589        ----------V-------R----------------T----T----.S.......-N-SD---..DNM-------------KF-YT-L-----------------------S--D--Q------I----    691
B.TW.TWCYS         ----------V-------R----------------N-N----I---.-.......-TYDE---..-NM---Q-E---D---DH-YN-L-K----------D-------------D--------RI----    679
B.UA.UKR1216       ----------V-------------------------T----.-.......---DK---..DNM-----E-------ID--YN-----------------A----------D-SK------I----    683
B.US.DH123         ----------V-------Q----------------T----T---.-.......---DT---..GNM---Q-EK------G--YN-----------------A----------S-------I----    681
B.US.JRCSF         ----------V----------M----------------T---.-.......---DS---..-NM-----EK--E---NT-YT------I-----------------------G--K------I----    680
B.US.JRFL          ----------V-------G-------------------------.-.......---DR---..-NM-----E---D----E-YT-------------------------D--K------I----    679
B.US.MNCG          ----------V-------------F-----------T--------.-.......---DD---..-NM---Q-E---D------Y--L-K--T---------------------D--------I----    689
B.US.RF            ----------V-------R----------------T--------.-.......---NM---..-NM---Q-E---D---GI-YN-L---------------------N-----D--Q-----RI----    697
B.US.SF2           ----------V-------R-------------------------.-.......----D---..DNM---Q-E---D---NT-YT-L-----------------------S---------I----    687
CONSENSUS C        -------qt-V--I--y-k----------------a-p--s---?-.-.......ksqee---..nnm---q--r-is--tnt-Yr-L-d--n--ek--kD--A--k-qn--s--d--n-----kI--m-    657
C.BI.BU910112      --------T-V--I-------------------------D---N-.-.......-T-GD---..-S----Q-----S---NT-YR-L-D------Q--KD--A----QN--S--D---------I----    673
C.BR.92BR025       --------T-V--I----R-------------------S---.-.......R-Q-D---..-NM---Q-----S---NT-YR-L-D----------D--A----QN--T--G---------I--K-    681
C.BW.96BW0402      --------T-V--I-----------------------N----N---.-.......---DY---..GNM---------S--SNT-YR-L-D---------KD--A--S-NN-----G--K------I----    684
C.CN.AF268277      --------T-V--I-------------------------H--S---.-.......R-Q-E---..-NM---Q-----S---NT-YR-L-D------R--KD--A--N-KN--S--D---------RI----    678
C.DJ.DJ373A        --------T-V--I-------------------------S---.-.......--Q-E---..DNM---Q----------ET-YR-L-V--T---Q--KD--A----QN--S--D--K------I----    675
C.ET.ETH2220       --------T-V--I--H-R-------------------S---.-.......--Q-E---..DNM---Q-----S---DI-YN-L-V-----D---KD--A----EN--------------I----    676
C.IN.93IN101       --------T-V--I-------------------------S---.-.......-TQSE---..-NM---Q----VS---NI-Y--L----------KD--A--S-KN--S--D---------I----    689
C.SO.SO145A        -------R--V--I----E-------------------S---.-.......RTQ-E---..ENL---Q--K--S---DT-YR-L-V---------KD--A--S-G------D--K------I--I-    686
C.UG.UG268A2       --------T-V--I----Q-------------------S---.-.......---GD---..DNM---Q-----S---GT-YR-L-D---------KD--A----QN--S--D--K------I----    681
CONSENSUS D        ----------?l---r--k------i-----rH--t-t----s---.-.......k-lde---..?NM--me-Er--d---Gl-Ys---e--t------ke--e----------s--q-----ki--mi    658
D.CD.84ZR085       ------------------------------H----T----S---.-.......R-V-Y---..GNM---Q-E---D---G--YN------I------K----------S--Q------I----    688
D.CD.ELI           ------------------------------H----N----S---.-.......R--NE---..QNM-----E---D---G--Y-------T------K----------S--Q------I----    686
D.CD.JY1           ----------V----S--------------H----T----S---.-.......----E---..-NM--I--E---D---GV-Y----N--I-------D--Q----------S--K------I----    695
D.CD.NDK           ----------V-------R----------RH----N----S---.-.......R--DE---..QNM-----E---D---G--Y-------I------K----------S--K----------    678
D.CI.CI13          ----------V----------------RH----T----S---.-.......---DE---..ENM-----E---D---G--Y-------T------K---A----------D--Q------I---V    673
D.SN.SE365A2       ----------------R----------RH----N----S---.-.......---TE---..QNM-----E---D---G--Y-------T----------A----------D--------I---V    683
D.TZ.87TZ4622      ----------XP---S--R----------RH----T--------.-.......---GE---..GNM-----E---D---G--Y-------T------D--Q----------S--Q------I----    683
D.UG.92UG024-D     ----------V----S----------RH--P-Q----S---.-.......---DT---..GNM-----E---S---G--YD------I------KD----------D--------R---I-    682
D.UG.WHO15-474     ----------------Q------S-----RH----T----S---.-.......---DD---..KNM-----EK--D---GV-YR------T----------Q----------Q------I----    692
F1.BE.VI850        ----------V-----------------------N----S---.-.......--Q-E---..-NM-----EK--S--SNI-YK----------------A----------D-S-------I----    664
F1.BR.93BR020-1    ----------V----------L----------N----S---.-.......-----E---..GNM-----EK-VS--SKE-YR---D----------A----------D--Q------I----    678
F1.BR.BZ126        ----------V-------Q------L----------N----S---.-.......--Q-E---..ENM-----EK--S--SYE-YR---Q----------A----------DMS------I----    674
F1.FI.FIN9363      ----------V----------L----------N----S---.-.......--QDE---..-NM---Q-EK--S--SKT-YM---K--S---R----------D---S--D---------I----    674
F1.FR.MP411        ----------V-----------------------N----T---.-.......--HDE---..-NM---Q-EK-----SNT-YR----------------A--------S--D-S-------I----    675
F2.CM.CA20         ----------L----------------I----T----T---.-.......-TQ-E---..KNM--IQ-EK--S---DT-YR---GA---------N--A----DI--S--S---------RI---V    663
F2.CM.HIM277819    ----------L----------F--L----------N----S---.-.......--QDE---..-NM---Q-EK--S---GT-YR---VA-----Q------A----DN-----T---------I----    679
F2.CM.MP255        ------------------------------------N-R--S---.-.......--YDD---..DNM---Q-EK--D---KT-Y----DA-----R------A----D---S--S---------I----    675
F2.CM.MP257        ------------------------------------P-T----L---.-.......--QDE---..GNM-----EK--G---DT-YR---SA---------D--A----DN-----S--R-----EI----    682
```

```
                                  immunodominant region
                                  |-      *     * -|   ^^^   ^        ^^       ^^^          ^^^                                ^^^
B.FR.HXB2                 VWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWS.N.........KSLEQIW..NHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMI    688
CONSENSUS G               -------qa-v--v--y-k-----------------N----t---.-........ksyne--..dnM--ie-e--is---qq-Ys-i-e--n-------D--A--k-as--?--?-tn-----kI-i-i    650
G.BE.DRCBL                -------R--V--L---------------------N----T---.-........--YNE--..ENM--I--E---D---YH-Y----Q--I-------D--A--Q-----S--S-S------RI-V--    679
G.FI.HH8793-12-1          ----------V--L----R-----------------N----------.-.....-TYND--..DNM--IQ-----S---QQ-Y---------------D--A--N-----T--D--K------I----    675
G.GA.LBV217               ----------V-------Q-----------------N----T---.-........--FNE--..DNM--I--E-------HQ-Y--L----I-------D--A----E---S--S--Q------I----    689
G.NG.92NG083              --------S-V--I----------------------N----T---.-........--YNE--..DNM--L--E---H---QH-Y---------------D--A-----------D-S------RI----    674
G.NG.NG1937               ----------V-----F----------------------N----T---.-.....R-YDE--..DNM--I--E--VS---QE-YK-------------D--A-----G-----S--K------I---V    695
G.NG.NG1939               ----------V----------------------------N----I---.-........--Y-D--..DKM--I-----VS---QQ-Y---------------D--A--------S--D---------I----    678
G.SE.SE6165               ----------L----------------------------N----V---.-........--YNE--..DNM--I--E-------YQ-Y--L------------D--A--Q--------G--R------I----    691
H.BE.VI991                ----------V----------------------------N----S---.-........---DE--..DNM------KQ-----DE-YR-L-V----------D--A-----N-----S---------RI----    694
H.BE.VI997                ---V------V----------------------------N----ST--.-........---AE--..DNM-------Q-D---EV-YR-L-L--T---Q---D--A----D------S---------I--I-    687
H.CF.90CF056              ----------V-------R------------------N----S---.-........--QSE--..DNM------KQ-S---EE-YR-L-V--T-------D--A--------T--D-SH------I----    680
J.SE.SE7022               ----------V----------------------------N----------.-.....--Y-D--..ENM--IQ-E-------GI-Y-----A-----T--KD--A----TN-------S-------I----    685
J.SE.SE7887               ----------V----------------------------N----------.-.....--Y-D--..ENM--IQ-E-------GI-Y-----A-----N--KD--A----TN-------S-------I----    683
K.CD.EQTB11C              -------R--V-------R------------------N----S---.-........--QSE--..ENM---Q-EK--S-H--T-YR------I-------D--A-----------D-S------I----    683
K.CM.MP535                -------R----------------------------N----S---.-........--W-E--..-NM-----EK--G--SDT-YK------T-------D--A-----------D--K------I----    674
N.CM.YBF106               I------R-KV--I----R---I-SL------T--Y-T----DX--S-.-.....T-YDT--..XNL--QQ---KVR--SGV-FD---QA-E--NT--KA-----Q--------D--K------IA--V    660
N.CM.YBF30                I---------KV--I----R---I-SL------T--Y-T----ET--N-.-.....T-YDT--..-NL--QQ--EKVR--SGV-FG---QA-E--NT--KS-----Q-D---S--G--K------IA---    677
CONSENSUS O               v---R--R--Ll-L-TLiQN--l-nL--cK-rli-Y-S-k--?t-???.....????nds--..dnl--Qe--qqi?-iSst-ydE-QkA-v---q-ekkl----E---iw--lD--K------iA-I-    649
O.CM.ANT70                X---R--R--L--L-TL-QN----SL---K---V-Y-S-K--RT-I.G.........N-S--..DTL--Q----Q-S-IS-T-YEE-QKA-V---Q--KK-----E---I---LD--K------IA-I-    684
O.CM.CM4974               ----R--R--LQ-L-TLIQN--R-NL---K-R---Y-S-K--TT-G.-.....YNDS--..DNY--QQ--QQV--VS-I-YDE-QKA-D---R-VKAV----E-------LD--K------IA-I-    694
O.CM.HIV1CA9EN            ----R--R--L--L-TLIQN----NL--YK-R---Y-S-K--DT-YK-.....TTDD--..DKL--Q---QQ---VS-I-YEE-QKA-V---E--KK-----E---IG--PD--K------MA-I-    697
O.CM.MVP5180              ----R--R--LQ-L-TLIQN--R-NL---K-----Y-S-K--T----.G.....RYNDDS--..DNL--QQ--QH---VS-I-YDE-QAA-D-----VKA-----E--------D--K------IA-I-    697
O.GA.VI686                ----R--R--L--L-TLIQN----NL---K-R---Y-S-Q--KT-TNK.....SNVD--N---..GNL--Q---QQ-D-IS-T-YDE-QKA-V---Q---K-----E---I---LD--K------MA-I-    700
O.GQ.193HA                ----R--R--L--L-TLIQN----NL---K--IV-Y-S-E--KT-I.G.........NDS--..-DL--Q---QQ-S-ISTT-FEE-QKA-V---Q--KK-----E---I---LD--K------IA-I-    681
O.SN.MP1300               ----R--R--L--L-TLIQN----NL---K-R-V-Y-S-K--RT-TN-.....NTD-DT---..GNL--Q---QQ-S-ISAT-YDE-QKA-V---H--KK-----E---I---LD--K------IA-I-    696
CPZ.US.CPZUS              ----------V-----------I--L------T--Y-T----DT--.-.....NL-YDA---..GNL--Q----KVR--SGT-F-----QA-E--NT--KS-----Q-S------D---------I-LIV    661
CONSENSUS 01              ----------V----------kF--L------Ii---a----s?--.-.....k-yee---..nnM--ie-er--?n--nq-YE?LTe--n--?r--kD--e--k-as-----?-tn-----kI--mi    662
01_AE.CF.90CF11697        ----------V----------KF--L------I----N----T---.-.....--YAE---..DNM--I---K------NQ-YE-LTK----------D--A-----N-----S---------I---V    688
01_AE.CF.90CF402          ----------V----------KF--L------I----S----S---.-.....--Y-A---..-NM--I--------E--NQ-YE-LT---D---R--KD--A----------D--R------I--I-    696
01_AE.CF.90CF4071         ----------V----------F--L------I----S----S---.-.....--YNE---..-NM--I----------NQ-YE-LTK--D---R--KD-------T------D--R-----RI----    683
01_AE.CM.CA10             ----------V----------F--L------I----S----S---.-.....--Y-E---..ENM--I--E-------KQ-YE-LT-------R--KD--A--S--------S---------RI----    694
01_AE.TH.93TH065          ----------V----------KF--L------I---------ST--.-.....R-Y-E---..-NM--T--E---S---NQ-YEILT---D--DR--KD--------N-----S-S------I----    679
01_AE.TH.CM240            ----------V----------KF--L------I---------ST--.-.....R-F-E---..-NM--I--E---S---NQ-YEILT------DR--KD----------D---------I----    682
01_AE.TH.KH03             ----------V----------KF--L------I---------ST--.-.....R-Y-E---..DNM--I--E---S---NQ-YEILT------DR--KD----------S---------I----    683
01_AE.TH.TH022            ----------V----------KF--L------IV--------ST--.-.....--Y-E---..DKM--TQ-E---S----T-YEILT------DR--KD--K----------D-SR-----RI----    682
CONSENSUS 02              ----------V--L-r--r-------------l----t----s---?-.-.....-Tynd---..dNM--lQ-dK-iS---di-Yn-i--s--q-----qD--A--k-as--n--d-tn--w--?I-imi    667
02_AG.FR.DJ263            ----------V--L-S--R-----------------N----S---.-.....-TYND---..DNM--LQ--K-VS---DI-YN-------------D--A--Q--N-----G--K------I----    694
02_AG.FR.DJ264            ----------V--L----R-----------------T---------.-.....-TYND---..DNM--LQ--K--S---DI-YN-------------KD--A----E---------K------I----    680
02_AG.NG.IBNG             ----------V--L----R-----------------T----S---.-.....-TFND---..DNM--IQ-EK--S---DI-YN-------R------D--A----------D---------RI----    679
02_AG.NG.NG1921           ----------V--L-G--R-------------I----------S---.-.....-TY-Y---..DNM--QQ--K-VS---NT-YR-L------------D--A--S-----S--D-S---R--QI----    698
02_AG.SE.SE7812           ----------V--L-A----------------------T----S---.-.....-TYDH---..GNM--LQ--K--S---HI-YD-------------D--A--------D--S-SS-----RI--I-    691
02_AG.SN.MP1211           ----------V--L----------------------T----S---N-.-.....-TFKD---..DNM--LQ--K--S---EK-YT----A---------D--A----------D--------LI-L-V    686
03_AB.RU.KAL153-2         ----------V----------------------------T---.-.....-P-DE-$..-NM-----E-------G--YN---------------I-A----------D-SK------I----    671
03_AB.RU.KAL68-1          ----------V--A-------------------------T---.-.....-P-DE---..-NM-----E-------G--YN---------------I-A--------S--D-SK------I----    671
03_AB.RU.RU98001          ----------V----------------------------T---.-.....---DK---..-NM-----E-------G--YN----------------A----------D-SK------I----    670
04_cpx.CY.94CY032-3       ----------V--L-S----------------------N----S---.-.....--YND---..DNM--LQ--K------QI-YG-L----------D--A----------S--K------I----    686
04_cpx.GR.97PVCH          ----------V--L-S----------F-------N----S--N.-.....--YND---..GNM--LQ--K------QI-YE-L----I-------D--AF----N-------S-------I----    707
04_cpx.GR.97PVMY          ---V------V--L-S--R----------R-----N----S---.-.....--Y-K---..DNM--LQ--K------EI-YD-LG-A---------D--A-----N--S--D-SH------I----    693
05_DF.BE.VI1310           ----------V----------L--S---H--S-S----S---.-.....--EGE---..DNM-----EK--S--S-T-YR---Q--I----------S--Q--------D--K------I----    693
05_DF.BE.VI961            ----------V-------R------L----------S----S---.-.....--Q-E---..ENM---Q-EK--GQ-SNT-YR---Q---------KD--A----------S-S------RI----    685
06_cpx.AU.BFP90           ------------------------------------P-N-L--T---.-.....-TYDE---..GNM--I---------QQ-Y----L--T-------D--A-----N--S--D---------I----    691
06_cpx.ML.95ML127         ------------------------------------P-N---------.-.....-TYNE---..DNM--I----------QQ-Y---------------D--A--------S--D-S------RI-VI-    677
06_cpx.ML.95ML84          ----------V----------------------P-N---------.-.....RTFNE---..-NM--I----------QQ-Y-------S-------D--A----------D-S-------I----    672
06_cpx.NG.NG3670a         ----------V--L----R------L--Y------P-N---------.-.....-TYNE---..DNM--IQ-----Y---QQ-Y-------S-------D--A----T---S--D-S-------I----    672
06_cpx.SN.97SE1078        ----------V--L----------------------P-N----I---.-.....-TYSE---..DNM--I----------QQ-YN-----------------A--------S--D-S-------I----    678
10_CD.BFL061              ----------V----S---------------RH----N----S---.-.....---GE---..DNM-----E---D---G--YN------I----------Q--------S--S---------I----    672
10_CD.BFL071              ----------------S----------------------N----D---.-.....----DA---...NM---D-E---D---G--Y--------------K----------------S---------RI----    683
10_CD.BFL110              ----------V----S----------------H---------S---.-.....----E---..-NM-----EK--D---G--Y-------T----------Q----------S--K------I----    680
CONSENSUS 11_cpx          ----------V--v-r--k-q----------------------n----v---?-.-.....-syde--??dNM--ie-e-e-d---qt-YT-l----n-------d--?--------n--d-sn-----?I---i    646
11_cpx.CM.MP818           ----------V----------------------------N----I---.-.....--YDE---..DNM--I--E---D---QT-YT-L----T-------D--S--------S--D---------I----    677
11_cpx.CM.CA1             ----------V----------------------------T----N---H-.-.....--VDD---..QNM--L--E---S---QT-YT-L----------------S--------S--E-S------RI----    680
11_cpx.FR.MP1298          ----------V----S--R----------------------N----I---K-.-.....--YNE---..-NM--IQ---------Q--YT-L----T----------S----------D-SS------I---V    687
11_cpx.FR.MP1307          ----------V--L----R-R----------------N----V---.-.....--YDE---..DNM--IQ-----H---QT-YT-L----S-------D--A--------D--S-S------RI----    682
11_cpx.GR.GR17            ----------V----------------------------N----V---.-.....-TYNE--ENENM--IK-E---D---QT-YT----------D--A----------D-S------RI----    676
CPZ.CM.CAM3               ----------V-------R---I--L------A--Y-T----NT--A-.-.....T-FDE---..-NL--QD--KRVK--SGV-F----QA-E--NT--KS-----Q-S------D--R-----------    681
CPZ.CD.CPZANT             ---V------M----K--R-----SL---AD-VT-H-T----N--VNFTQTCAKNSSDIQC---..ENM--Q----LVQ-S-GQ-YNILQIAHE---R-KK--Y-----S------D--Q------I----    690
CPZ.CM.CAM5               ----------V----------I-SL------A--Y-T----T--A-.-.....T-YDE---..-NL--QD--KKVK--SGV-F----QA-E--NT--KD-----Q-S---S-----Q------I-LIV    687
CPZ.GA.CPZGAB             I--V------L-------Q---I--L------AV-Y-T----N--P.G.....SN-TDD---..GNL--QQ--KLVS---GK-FG-L--A-S------RD-----Q--------D--K------I-L-A    679
```

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins

```
                          ^^^                 *
B.FR.HXB2                 AVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL$
CONSENSUS A               -in-lDti-iav-GWt-----ig-rig--iln--r-----f-rA-l-
A.GB.MA246                -IN--DTI--II-GW------IG-RF-----LN--------F--A---
A.KE.K89                  -SN-VDTI--V--GW------IA--IG---L---------F--A---
A.KE.Q23                  -IN-VDTI-----GW------IA-RIG---L---V--------A---
A.RW.PVPI                 -IN-FDTI--V--GW-------G-RIG--FLN--------A--A-I-
A.SE.UGSE8131             -I--FDTI-----GW------IG-RIG---L---------F--A---
A.UA.ukr970063            -IN-IDTI-----GW------IG-RF-----N--X-----A-KA-Q-
A.UG.92UG037              -IN--DTI-----GW------T--RLG---LN--------F--A---
A.UG.U455                 -IT--D-V-V---GWI-----IG-TIG---LN----------A---
A2.CD.97CDKS10            -I----S-------W------IG-R-----LN--------F--A---
A2.CD.97CDKTB48           -IR--DTI-V----W------IG-R------N----------A---
A2.CY.94CY017-41          -I--FDTI-V----W------IG-R-F---LN----------A---
CONSENSUS B               avsllnat-i--a-----vievvqRay-aiLh-prri---le-a-l-
B.AU.MBC18                ----F---------------AL-RTY---L----------A---
B.CN.RL42                 -IG---------------------R-Y---L---T--------A---
B.DE.D31                  ------------------------R-W---L---V--------A---
B.DE.HAN                  ----F-TI----------------R-----L-----V------A---
B.ES.89SP061              ----------------------L-R-----L----------A-Q-
B.GA.OYI                  VI--------------------I--R-Y--FLN----------A---
B.GB.CAM1                 ----FDTI----------------R-----L----------L---
B.GB.WB                   ----F-T--V-------------HRIW-G-L----------A---
B.JP.ETR                  ----------------------IL-RTY---L---VK-------A---
B.JP.JH32                 ----F--I-----------LKIL-R-F---L---T--------A---
B.KR.WK                   ---------V--------I--IL-R-Y---LN----------A---
B.NL.3202A21              ------------------------R----VL---V--------A---
B.TH.93TH067              ------T-----------L----R-Y---L---T-----FK-A---
B.TT.QZ4589               -------I----------I-KA--RI----L---------F--A---
B.TW.TWCYS                ----F--I----------------RVF---L---T--------A---
B.UA.UKR1216              ---------V-------------RR-F--FLR--------F--A---
B.US.DH123                ------------G-----I--IL-R-G---LN--T--------A---
B.US.JRCSF                ----------------I-----RVY---L---T--------A---
B.US.JRFL                 ----------------I--AL-RTY---L---T--------A---
B.US.MNCG                 ----------------------L-R-G---L---T--------A---
B.US.RF                   ------T----------I---A-RIL--FL-----------A---
B.US.SF2                  ---W--------T---------A-R-Y---L--H---------L---
CONSENSUS C               -is-ldtt--a-a-----i--?i-ric-aicn--r-----f-AA-q-
C.BI.BU910112             -IN---I----------I--II-RI----Y---------F-AA-Q-
C.BR.92BR025              -I--FDTI---------I---I--IW---CN--------F-AA-Q-
C.BW.96BW0402             -I---DT----------I--IA-RI----CN--T-----F-AA-Q-
C.CN.AF268277             ------------------------Y---L---------F-AA-Q-
C.DJ.DJ373A               -----DT----------I--II-RIW--FCN--------F-AA-Q-
C.ET.ETH2220              -IN---T---V-G-----F--LI-RIW--FCN----------AA-Q-
C.IN.93IN101              -I--FDSI--V-------I--L---F-----N--T-----F-AA-Q-
C.SO.SO145A               -I---DTI--T-------I--LA-RI--GV-N--T-----F-AA---
C.UG.UG268A2              -I---DTI----S-----I---G--IG---L---------F-AA-Q-
CONSENSUS D               ai--fdat-ia---g---iie?vqRacr-vlni-t-i--g---a---
D.CD.84ZR085              -I--VD-----------I-DI-RR--K-VL---T--------A---
D.CD.ELI                  -S--FD-I----------II-R----VLN----------S---
D.CD.JY1                  FI---------------I--LIRR-F--VL-----V------A---
D.CD.NDK                  -S---DTI------R---------R-----LNV---------L---
D.CI.CI13                 -I--FD---V--------IL-A--RVW--VL---T--------A---
D.SN.SE365A2              -I---DTI----------I-D---R-----L---T--------A---
D.TZ.87TZ4622             -I--F-T----------DI--RTV--VLN--T----S---A---
D.UG.92UG024-D            -I--F-T---V-------I--L--RIG---LN--T--------L---
D.UG.WHO15-474            -I--F-----T-------I---G-R------N----------A---
F1.BE.VI850               -I--F-T---V-------I---L-R-G--VLN--------A--A---
F1.BR.93BR020-1           -I----------W------AL-R-G---LN----------A---
F1.BR.BZ126               -I----T---V-T-----I-AAL-RIG---LN--T-----F--A---
F1.FI.FIN9363             -I--F-T---V----------AL-R-V--VLN-------RV--A-I-
F1.FR.MP411               -I----T---V----------L-R-G--VLNV---------S---
F2.CM.CA20                VI----T---V-------IV--L-R-G--VL---------F--A---
F2.CM.HIM277819           -I---D-----------I--II-RTF---LN----------A---
F2.CM.MP255               -I--FDTI----------I---L-R-G--VL---------A--F---
F2.CM.MP257               -I---DR----------I--IL-R-G--VLN----------A---
```

```
856
837
858
855
853
866
860
861
857
855
861
850
865
843
864
845
850
855
862
855
856
859
861
867
865
856
837
859
847
851
849
848
847
856
865
855
830
848
856
859
846
843
851
857
861
856
825
856
854
863
846
841
851
851
850
860
832
846
842
842
843
831
847
843
850
```

```
                          ^^^                 *
B.FR.HXB2                 AVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL$
CONSENSUS G               -In-lDTi---v-Nw------va-R-c--iLNi-r-i---l--A-l-
G.BE.DRCBL                -IN--DTI-----NW-------A-R-G--VLN----------A---
G.FI.HH8793-12-1          -IN--DTI-----NW------I--R-F--FLN--T--------A---
G.GA.LBV217               -IN--DTV---T-NW-------A-R-----LN----------A---
G.NG.92NG083              -IN--DTI---T-N--------A-R-Y---LNV-T--------A---
G.NG.NG1937               -IN-FDTI-----NW-------A-R-----LN--------A--A-Q-
G.NG.NG1939               -IN-FDTI-----NW-------A-R-R---LNV---V------A---
G.SE.SE6165               -I---DTV-----NW-------A-R-----LN--T--------A---
H.BE.VI991                -I----T----------I--L--R-W---L---------F--A---
H.BE.VI997                -IN---T---V-------I--I--R-W--VL----------------
H.CF.90CF056              -ID---T----------GI-VI--R-W---L---------F--S---
J.SE.SE7022               -I----T----------I--IA-R-F---L----------A---
J.SE.SE7887               -I----T----------I--IA-R-F---L----------A---
K.CD.EQTB11C              -IN---T----------I--I-YR-F--LL---------F--L---
K.CM.MP535                -I----T------G----I--IG-R-F--LL----------A---
N.CM.YBF106               -I----T---V-------F--LA-RIG-G-L----------A---
N.CM.YBF30                -I----T---V-------I--LA-RIG-G-L----------A-I-
CONSENSUS O               -ts-lDt?-Va--NW--gI-lGi-RIGrGiln--r-----l--?-l-
O.CM.ANT70                -T---DTL-V---NW--GI-AGI-RIGTG--N----------S---
O.CM.CM4974               -IT-FDSI-VS--NW--GI-LGI-RIGQG-LN--------A--F---
O.CM.HIV1CA9EN            -T---DTF-V---NW--SI-SGI-RIG-G-LN----------G---
O.CM.MVP5180              -TN--DTI-VS--NW--GI-LGL-RIGQGFL---------A----V-
O.GA.VI686                -T--IDTL-VT--NW--GI-LGI-RIG-G--N--T--------S---
O.GQ.193HA                -T---DTI-V---NW--GI-SGI-RIG-G-WN----------------
O.SN.MP1300               -T---DTL-V---NW--GI-LGL-RIG-G-LN----------A---
CPZ.US.CPZUS              -T---DT----------I--LTRRLFLG-I----------S---
CONSENSUS 01              aIs-ld-t-iav-GW---v--va-?-w--il---------l--a---
01_AE.CF.90CF11697        -IT-FD-I-V---GW---------R-W--LI----------A---
01_AE.CF.90CF402          -IT--D----T--GW------I--R-W---L----------A---
01_AE.CF.90CF4071         -IT--D----T--GW---A--IA-R-----L----------A---
01_AE.CM.CA10             -I---D----T--GW------GA-R-Y--FI---------F--A---
01_AE.TH.93TH065          -I---D-------GW-------A---W---L----------A---
01_AE.TH.CM240            -I---D-----A-GW-------A---W---L----------T---
01_AE.TH.KH03             -I---D-------GW-------A---W--FL----------A---
01_AE.TH.TH022            VI--------V--GW-------A---W---L----------A---
CONSENSUS 02              -In--dTi-iav-NW---v--ig-RvG---rN--r-----l--a-l-
02_AG.FR.DJ263            -IN--DTI-L---NW------IG-RVG---LN----------A---
02_AG.FR.DJ264            -IN--DTI-----NW------IG-RVG----N--V--------A---
02_AG.NG.IBNG             -IN---TI--V--NW---A--IG-RVG----N--------F--A---
02_AG.NG.NG1921           -IN--DTI--FT-NW------IG-R-G----N----------S---
02_AG.SE.SE7812           -I---DTI--V--NW------L--R-G---LN--------F--A---
02_AG.SN.MP1211           -IN--DT--V---NW------I--RTG---CN----------A-Q-
03_AB.RU.KAL153-2         -IN-IGTI-----GW------IG-RF---M-N--------A-KA-Q-
03_AB.RU.KAL68-1          -IN-IGTI-----GW------IG-RF-----N--------A-KA-Q-
03_AB.RU.RU98001          -IN-I-TI-----GW------IG-RF-----N--------A-KA-Q-
04_cpx.CY.94CY032-3       -IN-F-T----------I--A--R-----CN----------A---
04_cpx.GR.97PVCH          -IN--DT----------IL-A--R------N--------F-KA---
04_cpx.GR.97PVMY          -I---DT--V-------I--A--RI----LN----------A---
05_DF.BE.VI1310           -I----T--VV----------AL-R-G---LN----------A---
05_DF.BE.VI961            -I----T--VV-------IL-AL-R-G--VLN----------A---
06_cpx.AU.BFP90           -I--FD-A-----NW----A----RIF--FLNV-------F--A---
06_cpx.ML.95ML127         -IN--DT------N-------I--R-F--VLN--T-----F--A---
06_cpx.ML.95ML84          -I--IDT------NW---------R-F--VLN--------A--A-I-
06_cpx.NG.NG3670a         -IN-VDT------NW-------G--IL--FLN----------A---
06_cpx.SN.97SE1078        -I---DT------NW------I--RVF--FLNV-------F--A---
10_CD.BFL061              -I---DTI--E--GW------I--R-V---LN--T--------A---
10_CD.BFL071              -I---DT--V--------A--I--R-V-TVIN--------A--AF--
10_CD.BFL110              -I----T---V-------A--I--R-V--VLN--T--------A---
CONSENSUS 11_cpx          -is-l-a----------I--i??rvL-ailh--------?--A---
11_cpx.CM.MP818           -IN---T----------I--I--RVL-G-L----------A---
11_cpx.CM.CA1             ----------------I---AHRVL--FL---------F--A---
11_cpx.FR.MP1298          -I--F------------I--IAHR-L--FL----------A---
11_cpx.FR.MP1307          -I--F------------I--I--RVL---L---------F--A---
11_cpx.GR.GR17            -I---------------I---AHR-L---LN--------F--A---
CPZ.CM.CAM3               -I---D-----------IL--A-IIG-G-L----------S---
CPZ.CD.CPZANT             IIKNIDRL--W-G-K--SILLAL-TIV-I--EV---------IA-N-
CPZ.CM.CAM5               -T---DT------N----I--F--IVG-G-L----------S---
CPZ.GA.CPZGAB             -I---D-----------I--AF-VTL-I--N----------A---
```

```
856
824
854
850
864
849
870
853
866
862
855
848
853
851
851
842
836
852
819
859
869
872
872
875
856
872
836
835
863
871
858
869
854
857
858
857
840
869
855
854
866
859
861
839
839
838
855
875
862
861
853
859
845
840
840
846
840
851
848
808
845
848
855
850
844
856
858
862
854
```


HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

myristoylation
N-terminal alpha-helix
MHC-1 downregulation + protein kinase recruitment
heart of CD4 binding site
acidic cluster
PKC
SH3 binding
polypurine tract
PAK binding

```
B.FR.HXB2          MGGKWSKSSVIGWPTVRERMRRAE.....PAADRVGAASRDLEKHGAITSSNTAATNAACAWLEAQ.EEE.EVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQ.DILD    111
CONSENSUS A        --g----ssivg-pev-ermrrtp?????.....??????p--kG---V-Q--dkh--v--s-i??nhpSca-l---?eee?e-----r---------y-g-?--sh---e----d--iy-rk--.E---    111
A.FR.HIV232956     ---------IV---DI------TP.....---TG---V-Q--D-Y--V----I..NHPS-------.-D-.------R---------F-G-L-----------D---Y-R---.E---    108
A.KE.Q23           ---------IV---EI-------P.....---PG---V-Q--D----V--K-I..NHPSY------.-D-.D-----R----------G---------K----D--VY-RK--.E---    109
A.SE.SE6594        ---------IV---EI-------P.....SA--PG---V-Q--A----V----I..N-PS-V-----.---.------R----------G---------K----D--VY-RK--.E---    109
A.SE.SE8538        -------R-KE--SE---KI-QTP.....---KG---V-Q--DR--------I..NHPS-T-----.-D-.------K----------G-F-----------D---Y-RK--.E---    110
A.SE.SE8891        ---------IV---K-----A-TP.....---KG---V-Q--D----V----I..NHPS---V---.---.------R----------G---------K----D---Y-RK--.E---    109
A.SE.UGSE8131      --N----.....-E----I-Q-RAPA.....HTPAPT--TG---V-Q----------I..NHPS-T-----AQ-DE-----R----------G-L---------Y-RK--.E---    109
A.UG.92UG037       --N-----CIV---E----I-QTPTAARE....RTRQAPT--KG---V-Q--D----V----V..NHPS-V-----.---.------R----------F--GF--------D---Y-KK--.E---    115
A.UG.U455          -------K-RVE---E--K---ETP.....A--KG---V-Q--D-Y--V-----SS---S-------.--G.D-----R----------F---F--------D------K--.E---    120
A2.CD.97CDKTB48    -------RTIV---EI------TPPAAE.....GVRPTP---EG---V-Q--AR---V-------N-PD-------.---.------R------A----G----------D---Y-----.----    111
A2.CY.94CY017-41   -------R-IP---AI------TPPTAQ.....RTEAVS---PG---V-Q--AT---V---------PD---V---.---S------R---------F-G-F---F--------D---Y--K--.----    121
CONSENSUS B        --gk-s-?s?g-wptv-e-mrrae?????????????????pa-dG---v-r--ekhgai-ss-taat--dcawl-aq?eee?ev----r-------mtykaal-l----kek---e-liy-qkrq?d---    122
B.AU.MBC200        -------C..A-----------APGER.....RRRAE----G---------R---L-------N--D-------.---G------R----------G-L------R---------Y-----.----    109
B.CN.RL42          -------H-MF---S-----K---.....---G---V-----------------.---.------R----------G-L------R---------Y--E--.----    119
B.DE.D31           ----------V---AI----K---.....---EG---V--------------P-----------.---.------K-----------------------V---K--.----    111
B.DE.HAN           -------C..S---------KQ--.....PE----G--------K----------TN----------.---.E----R----------G-L---------Y-PK--.E---    111
B.ES.89SP061       ---------K-----------.....---G---V----------------TN--D-------.---.------R----------G-L---------W--K--.----    111
B.FR.NE100         -------N-MS--T-I-----Q--.....---G---V----------------D-------.---.------R---------RR--I------R--------V---K--.E---    111
B.FR.SWB884        ---------MS---AI----K---.....---G---V----------------D-------.---.------R---------RR--I------R--------V---K--.E---    111
B.GA.OYI           -------C-MK----I----K---LQ.....PPE---EG-------------D-------.-D-.------R----------G-L---------Y--K--.----    116
B.GB.CAM1          -------R-LG--SA-----Q---P.....RAE---EG---V----------D-------.---.------R----------L-I---------I-Y-----.----    115
B.IN.HIVP35A       --------KMG---A----I--TDPAER.....IRQPE---AG---V-----R----------D-------.-D-.D-----R----------G-L------RK------I-Y--K--.----    120
B.IT.B-IT-R5       --NX-X---XS-X-A---------.....---G---V-----R------X---TN--D------H.-D-.------R-------X-CR--I---------VW--K--.----    111
B.JP.nef<7>-a      -------R-GE---K--------R.....PAE----G---V-----R------N----N--D-------.---.------R----------G---------V---Q--.----    114
B.KR.CSR9412d      -------R..---A-----E---.....---G---------Y----------N--DL------.---.------R----------G--V---K--.----    108
B.KR.WK            -------R--P--N-I-K-----.....---EG---------QR----T----SN------Q---.---.------R----------S-L---------VY--K--.----    111
B.NL.3202A21       ----------V---AI----K---.....---G---V----------D-------.-D-.------K----------G-L---------Y-----.----    111
B.SE.AF047085      -------C--V---------K---.....-P--G---V----G----------TN--D-------.-D-.N-----R----------F--------Q---------.----    111
B.TH.28-19         --A----F..G---A--------QPETRERRRPTQQPATE---VG---V----------E---D--R--S-.---.------K----------G-L---------K--.E---    125
B.TH.AF082839      -------R-GG--NA-------T-.....P--G---V-Q--G----------P-N--D-------.---.------R---------------------HK--.----    110
B.TW.TWCYS         -------R-IP--SNI---I-Q--P.....A---G---V----------D------Y.-D-.------R----------I---------RKE------VY--KKEX----    113
B.US.JRCSF         -------H--P--S----------PATD.....RVRQTE---VG---V----------D------Y.-D-.------R----------I---------Y--K--.----    121
B.US.JRFL          -------R--P--S----------PAAD.....RVRRTE---VG---V----------D-------.-D-.------R----------G---------K--.----    121
B.US.LM1           --S-----IRA--S----------.....---G---V----------A-HN--D-------.---.------R----------G-L--------E-------Y-----.----    111
B.US.MN            -------R.-T-------------P.....AEL---G----------L---------D-------.---.------K----------L---------D---Y--K--.----    113
B.US.NC7           ---------IV----I----K--G.....---G----------R----------D-------.---.------R----------GI--------------VW-----.----    111
B.US.RF            --------KMG---A-----QK--.....---G----------T--------N----T-----ED-DE------R---------F---------------D--VF--K--.----    113
B.US.SF2           -------R-MG--SAI--------P.....RAE----G---V----------------D-------.---.------R----------L-I---------------W-----.E---    115
CONSENSUS C        --gkw--csivgw-ai-e-mrrte???.....?????p--eG--aa-q-ldky--l-s-n-dtt-adcawleaq.eeeee-g---r---------y-gafd-sf--kek---egliY-kK--.e---    112
C.BR.92BR025       --N----C-TV-R-AI-------.....---EG--P--Q-SD-Y--L---S-P-N--D-------.Q--E------R----------V----F--E---------Y-KK--.----    111
C.BW.96BW01B21     ---------IV---A----I--T-.....---EG-----Q--D-Y--L-T---DPN-------R--.---G------R----------G-F--GF----------Y-KK--.E---    112
C.BW.96BW0402      ---------IV---A-------T-.....---EG--T--Q--D----L-T-----N----------.---E------R----------F----------I-Y-KK--.----    112
C.BW.96BW0502      -------C..S---A-------TRPAV.....EGRTES--EG-----Q--D-Y--L-----STN--D-------.---GD-----R----------G-F--GF----------VY-KK--.E---    118
C.BW.96BW1104      -------R-K-E---I-D----T-PAA.....EGVGA--EG-----Q--D----L-------N--D-------.---.------R----------FG--F----------Y-RK--.E---    118
C.ET.ETH2220       ---TM--C-PV---AI---I---A.....---EG--------D-Y--L-----P-N-PD-------.---E------R----------F---L----------Y-KK--.E---    112
C.FR.HIV232980     --N----C-PV---A-------TVPE.....AERAG--VG-----Q--D----L-----P-N--D-------.---E------R----------F--------D--TY--K--.----    118
C.FR.HIV232996     --S----C-PV----------QTN.....PA---EG-----Q--D-Y--L-T---P-N--D-------.---GG-----R----------G-F---F----------Y--K--.E---    114
C.IN.93IN101       -------C-IV---AI--------.....---EG-----Q--D----L-----DT---D----R--.---G----------------S-F---F----------Y-KK--.E---    112
C.IN.93IN904       -------C-IV---DI------TQ.....---EG-----Q--D-Y--L-----DI--PD----K--.---E------R---------F-E-----F----------Y-KK--.E---    112
C.IN.93IN999       -------C-IV---A-------T-.....---EG-----Q--D-Y--L-----DT---D----R--.---E------R---------F-G-F---F----------Y-KK--.E---    112
C.IN.94IN11246     -------C-IV---EI------TQ.....----G-----Q--D-Y--L-----DT---D----RT-.---E------R---------F-G-F---F----------Y-KK--.E---    112
C.IN.95IN21068     -------C-IV---DI------T-.....---EG-----Q--DRY--L-----DT--TE.....---.G-----R---------F-G-L---F----------Y-KK--.E---    104
C.IN.HIVY15117     -------C-IV---DI------T-.....---EG-----Q--D-----L-----DT---D.....-DG.--A---R----------G-F---F----E----A--Y-KK--.E---    104
C.IN.HIVY17884     ---------IV---AI------T-P.....AAE---EG-----Q--DR---L-----DT---D----R--.--DE------R---------F-G-F---F----------Y-KK--.E---    116
C.IN.HIVY17891     -------C-IV---AI------T-.....---EG-----Q--D----L-----DT---D----RT-.---E------R---------F-G-F---F----------Y-KK--.E---    112
C.IN.HIVY17892     ---------IV---AI----K--Q.....---EG-----Q--D-Y--L-----DT---D----R--.---EG-----R---------F-G-F---F----------Y-KK--.E---    112
D.CD.84ZR085       ---------IV---AI---I-KTDPRE.....RRRPE----G---V-----R---------T-------V---.--DE------R----------------RK-------VY--K--.----    120
D.CD.ELI           ---------IV---AI---I--TN.....---G---V----------S---D-------.--SD------R----------E-L---------W-KK--.E---    112
D.CD.NDK           ---------LV---AI---I-KTD.....---G---V----------S--DT-------.--SE------R----------E---------W-KK--.E---    112
D.UG.94UG1141      ---------IV---A-------T-.....---EG----------------Q--D-------.---.------R----------E---------VW-PK--.E---    111
F1.BE.VI850        ---------IV---A-G----QTP.....T--EG---V----DRR--------RT--PDL-------.---.------R----V---------Y-KK-G.-T--    111
F1.BR.93BR020-1    ---------IV---AI------TPP.....TP---EG---V-Q---RR--------R-N-PDL-------.--D.------R----------G---------Y-K---.E---    114
F1.FI.FIN9363      ---------IV---AI------PP.....PA--EG---V-Q---RR--------G---PDL-------.-D-.------R----------G-F---Q-XX--------XY-KK--.E---    112
F1.FR.MP411        ---------IV---A-------TP.....---EG---V-Q--DRR--V----------PDL-------.-D-.------R---------F---------------Y-KK--.E---    111
F2.CM.HIV232985    ---------IV---N-------TP.....V--EG---KV-Q--D----------R----DL-------.-D-.------R---------------------Y-KK--.E---    111
F2.CM.MP257        ---------IV---AI---I--TP.....V--EG---V-Q--D-R----N---G----DL-------.-D-.------R----------L---------Y-RK--.E---    111
```

```
                                                
```

HIV-1/SIVcpz proteins

```
                                                
```

```
                                                
```

HIV-1/SIVcpz proteins

EXHIBIT B

JOURNAL OF VIROLOGY, Oct. 2000, p. 9668–9679 Vol. 74, No. 20
0022-538X/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Comparison of Second-Strand Transfer Requirements and RNase H Cleavages Catalyzed by Human Immunodeficiency Virus Type 1 Reverse Transcriptase (RT) and E478Q RT

CHRISTINE SMITH SNYDER AND MONICA J. ROTH*

Department of Biochemistry, University of Medicine and Dentistry of New Jersey—Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Received 1 March 2000/Accepted 7 July 2000

Truncated tRNA-DNA mimics were examined in an in vitro assay for second-strand transfer during human immunodeficiency virus type 1 (HIV-1) reverse transcription. Strand transfer in this system requires the progressive degradation of the RNA within the 18-mer tRNA-DNA (plus-strand strong stop DNA) intermediate to products approximately 8 nucleotides in length. The ability of the truncated substrates to substitute for directional processing by RNase H or reverse transcriptase (RT) was examined. Using wild-type HIV-1 RT, substrates which truncated the 5′ end of the tRNA primer by 6, 9, and 12 nucleotides (Δ6, Δ9, and Δ12, respectively) were recognized by RNase H and resulted in strand transfer. An overlap of 5 nucleotides between the acceptor and newly synthesized DNA template was sufficient for strand transfer. The mutant RT, E478Q correctly catalyzed the initial cleavage of the 18-mer tRNA-DNA mimic in the presence of Mn^{2+}; however, no directional processing was observed. In contrast, no RNase H activity was observed with the Δ6, Δ9, and Δ12 substrates with E478Q RT in this strand transfer assay. However, when complemented with *Escherichia coli* RNase H, E478Q RT supported strand transfer with the truncated substrates. E478Q RT did cleave the truncated forms of the substrates, Δ6, Δ9, and Δ12, in a polymerase-independent assay. The size requirements of the substrates which were cleaved by the polymerase-independent RNase H activity of E478Q RT are defined.

Reverse transcription is a multistep process that is carried out by one virus-encoded enzyme, reverse transcriptase (RT). RT is a multifunctional enzyme which possesses RNA-dependent and DNA-dependent polymerase activities and RNase H activity. RNase H functions to remove RNA when it is present in an RNA-DNA hybrid (for a review, see reference 2).

In the process of reverse transcription, the viral RNA is converted to double-stranded DNA, which is subsequently integrated into the host genome. Human immunodeficiency virus type 1 (HIV-1) reverse transcription is initiated (3, 25–27, 33) using the cellular $tRNA^{Lys,3}$ as a primer (28, 35, 36). The first 18 3′ nucleotides of the tRNA primer are complementary to the primer binding site (PBS) sequence on the viral genome. Elongation of minus-strand synthesis pauses at the 5′ terminus of the viral RNA (31), completion of which requires a strand transfer, referred to as minus-strand or first-strand transfer. Plus-strand synthesis is initiated at the polypurine tract and continues through cDNA synthesis of the first 18 nucleotides of the tRNA primer (2, 4, 7). The second-strand transfer requires removal of the tRNA primer (4, 46, 50, 58), which allows an acceptor PBS molecule to enter, and subsequent completion of viral DNA synthesis.

HIV-1 RT consists of a heterodimer of two subunits, p66 and p51 (14, 34). The p66 subunit consists of the polymerase and the RNase H domains; the p51 subunit lacks the RNase H domain. Based on resemblance of the crystal structure to a right hand (1, 30), the subunits have been further divided into subdomains: palm, finger, thumb, connection, and RNase H. Mutagenesis studies have identified interactions between the polymerase and RNase H domains. Mutations within the thumb subdomain and the primer grip and deletions in the p51 C terminus decrease RNase H activity (9, 18, 21, 44).

During the viral life cycle, RNase H functions to degrade the viral genome, generate and remove the polypurine tract primer, and remove the tRNA primer. Removal of the tRNA primer has been extensively characterized for HIV-1 RT, Moloney murine leukemia virus RT, and an isolated HIV-1 RNase H domain (48, 49, 59). RNase H activity has been classified as either polymerase dependent or polymerase independent (2). The polymerization active site is spatially separated from the RNase H active site; polymerase-dependent RNase H activity results in RNase H cleavages which lag approximately 18 to 20 nucleotides behind the site of polymerization (55, 56). The catalytic residues of the RNase H active site are Asp 443, Glu 478, and Asp 498 (13, 17, 47). The requirements of RNase H activity during HIV-1 reverse transcription have been further characterized through the analysis of an RNase H-defective mutant, E478Q RT. This RNase H mutant possesses only Mn^{2+}-dependent RNase H activity. Additionally, it is capable of only a single endoribonucleolytic cleavage and lacks the ability to further degrade the tRNA primer (10).

Previously, we showed that RNase H activity is required for the HIV-1 second-strand transfer (50). More specifically, a single endoribonucleolytic cleavage is not sufficient to allow release of the tRNA primer. Rather, subsequent RNase H degradation must be carried out by RT or through complementation with *Escherichia coli* RNase H. We have now investigated the ability of HIV-1 RT and a mutant enzyme, E478Q RT, to support strand transfer with substrates possessing truncations in the 5′ portion of the tRNA primer. These substrates have the potential to substitute for RNase H-catalyzed directional processing, due to their decreased melting temperatures (T_m). If properly cleaved by E478Q RT, the truncated substrates could potentially support strand transfer without

* Corresponding author. Mailing address: Department of Biochemistry, University of Medicine and Dentistry of New Jersey—Robert Wood Johnson Medical School, Piscataway, NJ 08854. Phone: (732) 235-5048. Fax: (732) 235-4783. E-mail: Roth@waksman.rutgers.edu.

complementation by *E. coli* RNase H. In polymerase-independent assays, the truncated substrates were recognized and cleaved by E478Q RT. However, the results indicate a differential recognition by wild-type (WT) RT and E478Q RT on the truncated substrates in second-strand transfer reactions. Utilizing an in vitro strand transfer assay, we have shown that 5 nucleotides of overlap between the newly synthesized DNA strand and the acceptor molecule is sufficient for strand transfer.

MATERIALS AND METHODS

Enzymes and nucleotides. [γ-^{32}P]ATP was purchased from ICN. T4 polynucleotide kinase was purchased from New England Biolabs. RNasin was purchased from Promega. Klenow Exonuclease (−) was purchased from Boehringer Mannheim. HIV-1 RT was obtained either from Jeffrey Culp and Christine Debouch, Department of Protein Biochemistry, SmithKline Beecham Pharmaceuticals (37), or from Stuart Le Grice. The two HIV-1 RT preparations displayed equivalent specificity in the tRNA removal assay, indicating that variations in the expression and purification schemes did not result in altered biochemical properties (data not shown). The presence of the histidine tag on the HIV-1 RT has been previously shown not to influence catalytic functions (32). HIV-1 RT E478Q mutant was obtained from Stuart F. Le Grice and the AIDS repository (contributor, Stuart F. Le Grice). HIV-1-isolated RNase H (NY427) was purified from *E. coli* containing plasmid pET-NY427 (52). *E. coli* RNase H was purchased from Gibco BRL.

Oligonucleotides. The RNA-DNA hybrid oligonucleotides (RNA is indicated in bold) 17-mer (5′ **GUUCGGGCGCCACUGCT** 3′), 14-mer (5′ **CGGGCGCCA** CTGCT 3′), 11-mer (5′ **GCGCCACTGCT** 3′), 17-mer (DNA) (5′ AGCAGTG GCGCCCGAAC 3′), 14-mer (DNA) (5′ AGCAGTGGCGCCCG 3′), 11-mer (DNA) (5′ AGCAGTGGCGC 3′), HTD-1 (5′ GTGTGGAAAATCTCTAGCA GTGGCGCCCCGAACAGGGA 3′), 17080 (5′ ATCTCTAGCAGTGGCGCC CGAACAGGGAC 3′), and 17081 (5′ GAAAATCTCTAGCAGTGGCGCCC GAACAGGGAC 3′) were synthesized by Integrated DNA technologies. Oligonucleotides 5785 (5′ CCCTCAGCCCTTTTAGTCAGTGTGG 3′), 5786 (5′ CCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAAC AGGGACCTGAAAGCGA 3′), 5331 (5′ AGCAGTGGCGCCCGAACGCGG GGCTTGTCCCT 3′), and 5580 (5′ TTTCGCTTTCAGGTCCCTGTTCGGGC GCCA 3′) were synthesized by the University of Medicine and Dentistry of New Jersey.

Strand transfer substrate preparation. The truncated RNA-DNA hybrid strand transfer substrates were prepared as follows. Portions (20 pmol) of the 17-mer, 14-mer, and 11-mer RNA-DNA hybrids were 5′-end labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. The radiolabeled RNA-DNA oligonucleotides were isolated utilizing G-25 spin columns (Boehringer Mannheim). The labeled substrates were annealed to 40 pmol of oligonucleotide 5786 in a 25-μl reaction mixture containing 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 8 mM $MgCl_2$, and 2 mM dithiothreitol (DTT). The substrates were extended using Exonuclease (−) Klenow, isolated on gels, and eluted overnight as previously described (50). The product sizes for the input 17-mer, 14-mer, and 11-mer RNA-DNA hybrids are 44-mer, 41-mer, and 38-mer, respectively. The substrates were then annealed to oligonucleotide 5785 (referred to as 26-mer). This oligonucleotide was also 5′-end labeled in the same manner as described for the RNA-DNA hybrids.

Strand transfer reactions. Reactions were performed as previously described (50). Briefly, the annealed RNA-DNA hybrid substrates were incubated with either HIV-1 RT or E478Q RT. The strand transfer reactions were performed in a 20-μl reaction mixture containing approximately 4 pmol of substrate (substrate refers to the 50-mer RNA-DNA oligonucleotide annealed to primer 5785), 4 pmol of acceptor (oligonucleotide 5580), and 2 pmol of either HIV-1 RT or E478Q RT in a reaction buffer containing 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 8 mM $MgCl_2$ or 8 mM $MnCl_2$, 2 mM DTT, and 0.25 μM each deoxynucleoside triphosphate (dNTP) (dATP, dCTP, dGTP, and TTP). Reactions were initiated upon the addition of enzyme and performed at 37°C. Aliquots (2.5 μl) were removed at the indicated time points. The reaction products were analyzed on a 15% polyacrylamide denaturing gel and exposed to autoradiography film.

In reactions in which multiple enzymes were used, such as *E. coli* RNase H, 1 pmol of the additional enzyme was added after the 12-min time point.

RNase H cleavage assays. Reactions were performed as previously described (49). Briefly, approximately 4 pmol of the indicated RNA-DNA hybrid substrate was incubated with either 1 pmol of HIV-1 RT or E478Q RT. RNase H cleavage substrates were prepared in the same manner as described for the strand transfer substrates. The annealing templates used for extension templates were oligonucleotides 5786, HTD-1, 17081, 17080, and 5531 for substrates B to F (see Fig. 6), respectively. These newly extended substrates were gel isolated and eluted as previously described (50). The substrates were again annealed to their corresponding extension templates. The reaction mixture (20 μl) contained 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 8 mM $MnCl_2$, and 2 mM DTT. Time course reactions were performed, and 3-μl aliquots were removed at 0, 2, 5, 15, and 30 min. Samples were analyzed on a 20% polyacrylamide denaturing gel and exposed to autoradiography film.

RESULTS

Strand transfer assays with truncated substrates. The second-strand transfer is a vital intermediary step in the synthesis of full-length double-stranded retroviral DNA. An in vitro assay has been developed to analyze the second-strand transfer reaction during HIV-1 reverse transcription (50). In this assay, model substrates were constructed which mimic the intermediate formed during plus-strand strong stop synthesis in which the first 18 nucleotides of the tRNA primer are reverse transcribed (Fig. 1, step 2). In the removal of the tRNA, the initial cleavage occurs at the penultimate nucleotide at the 3′ end of the tRNA (step 3) (51, 53). Previous studies have indicated that the tRNA needs to undergo further degradation for strand-transfer to proceed (50). Once the tRNA is removed, the acceptor molecule can enter and a 70-mer strand transfer product results (step 4).

Previous analysis indicated the RNase H-defective mutant, E478Q RT, was capable of performing only a single endoribonucleolytic cleavage in removing the tRNA primer (10, 50). Using our model system, this cleavage yields a 17-mer RNase H cleavage product, which was not sufficient for strand transfer to proceed (50). Therefore, it was reasoned that decreasing the size of the RNA component would lower the T_m of the RNA-DNA hybrid and could compensate for the lack of directional processing by E478Q RT. Model substrates were constructed that truncated the 5′ terminus of the RNA by 6, 9, or 12 nucleotides and were named Δ6, Δ9, and Δ12, respectively (Fig. 1). These truncated substrates were specifically recognized and cleaved at the expected position by both WT HIV-1 RT and an isolated HIV-1 RNase H domain (49).

Figure 1 summarizes the second-strand transfer assay with substrates possessing truncations in the 5′ RNA portion. Step 1 represents the input substrates for the Δ6, Δ9, and Δ12 constructs, in which the RNA-DNA oligonucleotides are hybridized with plus-strand primer 5785. In the presence of dNTPs and WT HIV-1 RT, polymerization occurs (step 2) and the full-length plus-strand product is synthesized. In this complex, the RT is paused with the terminus of the nucleic acid substrate bound in the polymerase active site. Once the RNA-DNA hybrid is formed, RNase H can cleave the RNA primer (step 3). Using a full-length 18-mer RNA, at the completion of plus-strand synthesis, the E478Q RNase H active site is optimally positioned for the single endoribonucleolytic cleavage to occur through a polymerase-dependent mechanism. However, with the truncated substrates, a polymerase-independent RNase H cleavage would be required to remove the RNA, since in a polymerase-dependent assay, the RNase H active site would be positioned within a double-stranded DNA region. Once the RNA primer is removed, the acceptor molecule enters and a 70-mer strand transfer product is produced (step 4). Decreasing the size of the RNA template affects the overlap between the newly synthesized DNA strand and the acceptor molecule. As the truncation increases, the amount of overlap between the acceptor and newly synthesized strand decreases. This assay therefore allows the determination of the minimum overlap sequence required for strand transfer to occur.

HIV-1 RT assayed with truncated substrates. The truncated substrates (Δ6, Δ9, and Δ12) were assayed with WT HIV-1 RT (Fig. 2) in the presence of Mg^{2+}, dNTPs, and acceptor molecule. Reactions performed in the presence of Mn^{2+} yielded equivalent results (data not shown). Figure 2 represents time courses from 0 to 30 min. Lanes 1 to 6 represent HIV-1 strand

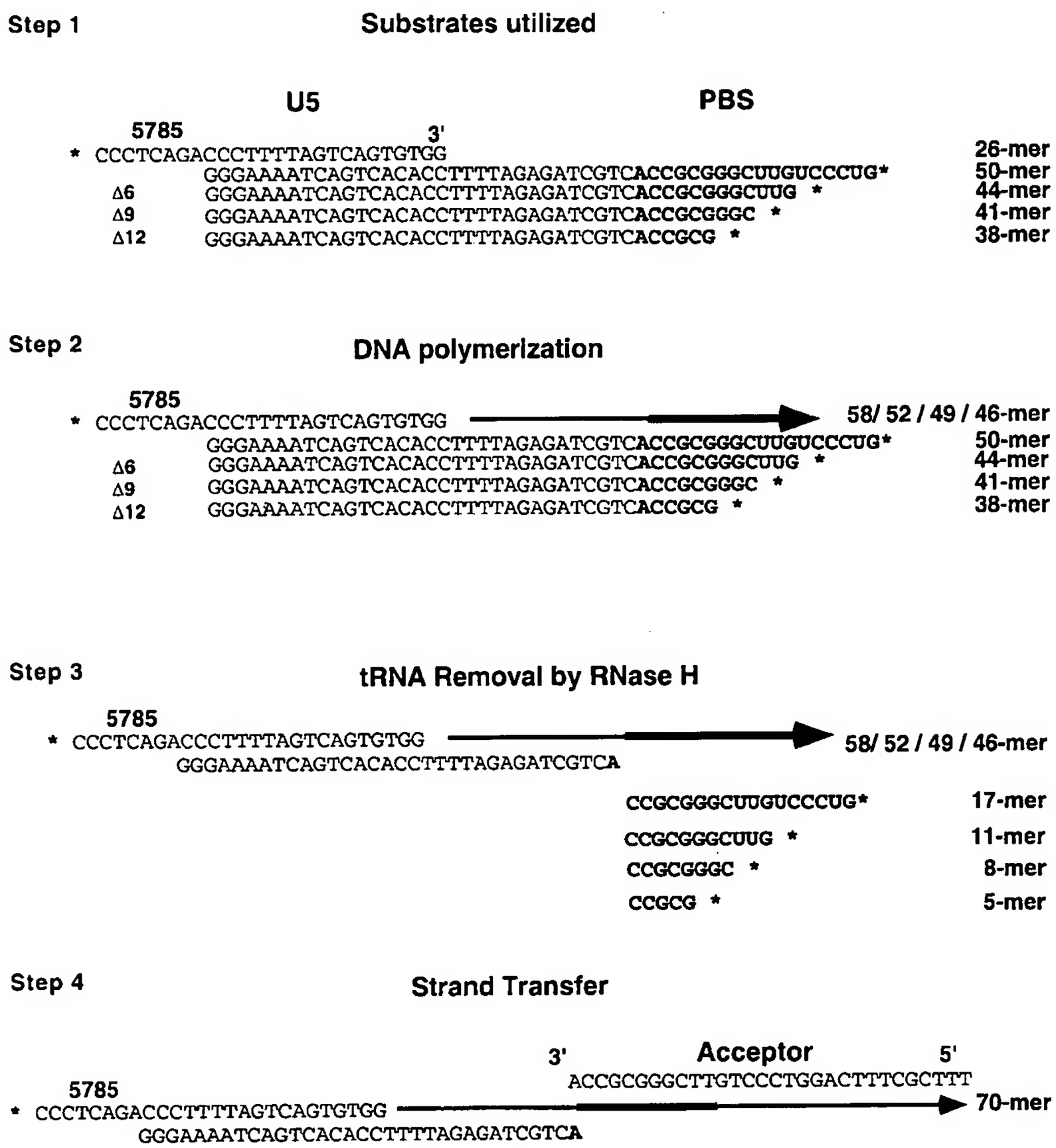


FIG. 1. Second-strand transfer assay with model substrates. This illustrates the model strand transfer assay with the truncated substrate possessing only 12 (Δ6), 9 (Δ9), and 6 (Δ12) positions of the RNA sequence. Step 1 illustrates the input substrate for each truncated substrate, along with their respective input RNA-DNA sizes. Step 2 illustrates the polymerization reaction which can occur in the presence of RT and dNTPs and the size of the polymerization product for each substrate, 52-mer (Δ6), 49-mer (Δ9), and 46-mer (Δ12). DNA polymerization creates the RNA-DNA hybrid, which is a substrate for the RNase H domain (step 3). Once the RNA has been removed between the terminal ribo-A and ribo-C, the acceptor molecule can enter and produce a strand transfer product, 70-mer (step 4). In each step, the RNA portion is indicated in bold and the 5′ radiolabel is indicated by an asterisk. The size of the strand transfer product (70-mer) would be the same for each truncated substrate.

transfer reactions with the Δ6 construct. The plus-strand oligonucleotide, 5785, was quickly extended from a 26-mer to the 52-mer product. The RNA portion of the RNA-DNA hybrid was degraded by the RNase H domain, and a strand transfer product (70-mer) was produced. Similarly, HIV-1 RT was capable of polymerization, RNase H activity, and strand transfer on the Δ9 substrate (lanes 7 to 12). The Δ12 construct was also assayed with HIV-1 RT (lanes 13 to 18). This construct had the largest deletion, possessing only 6 nucleotides of the tRNA primer. Therefore, there was only 5 bp of overlap between the acceptor and the newly synthesized DNA strand to support strand transfer. This substrate also successfully produced a strand transfer product (70-mer), albeit at lower efficiency. This result indicated that a 5-nucleotide overlap between the newly synthesized DNA strand and the acceptor strand is sufficient to support strand transfer.

E478Q RT assayed with the truncated substrates. The truncation substrates were assayed with E478Q RT in the same manner as for HIV-1 RT, except that reactions were performed in the presence of Mn^{2+}. Previous analysis indicated that the addition of Mn^{2+} to the WT HIV-1 RT does not change the RNase H or strand transfer properties of the enzyme, allowing direct comparison of the WT and E478Q mutant RT (50). It had been postulated that the single E478Q RNase H cleavage on the Δ9 and Δ12 substrates would release 8-mer and 5-mer RNA species, respectively. The reaction temperature is above the T_m of the products and would allow dissociation of the RNA. The T_m of the Δ6 construct, at ap-

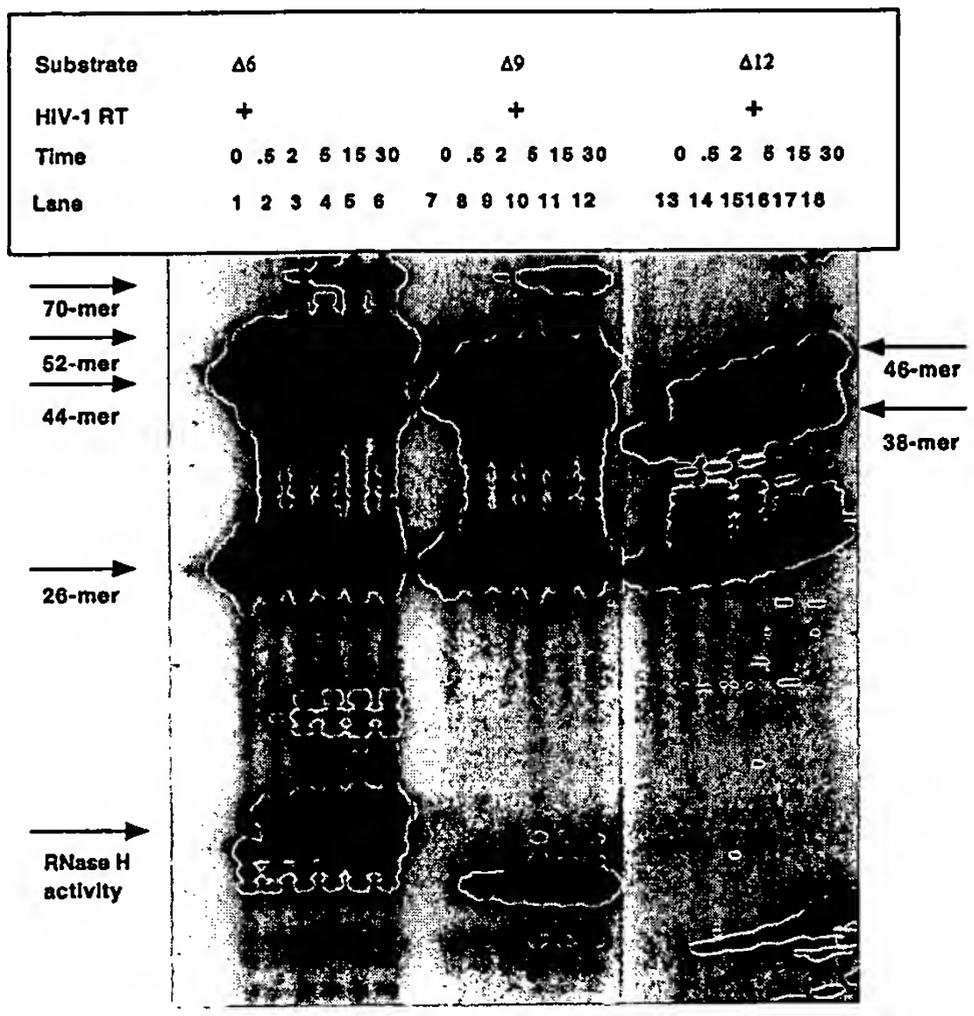


FIG. 2. Truncated substrates assayed with HIV-1 RT. Reactions were performed as described in Materials and Methods. Lanes 1 to 6, 7 to 12, and 13 to 18 represent HIV-1 RT incubated with the Δ6, Δ9, and Δ12 constructs, respectively. Time points are indicated above each lane in minutes. Strand transfer products (70-mer), DNA primer (26-mer), and RNase H products are indicated by arrows. Input substrates for the Δ6, Δ9, and Δ12 constructs are 44-mer, 41-mer, and 38-mer, respectively. Initial RNase H cleavage products for the Δ6, Δ9, and Δ12 constructs are 11-mer, 8-mer, and 5-mer, respectively.

proximately 42°C, may be too high for dissociation to occur. In previous studies, strand transfer correlated with the appearance of 8-mer RNA products (42, 50). Figure 3, lanes 1 to 6, lanes 7 to 12, and lanes 13 to 18 represent E478Q RT assayed with the truncations Δ6, Δ9, and Δ12 respectively. For all of the constructs, extension products were observed, indicative of complete synthesis. Interestingly, no RNase H cleavage products were observed for any of the constructs assayed in the presence of E478Q RT. Reactions performed for up to 40 min and/or in the presence of both divalent cations did not yield RNase H cleavage products (data not shown). A band was observed with the Δ6 construct (lanes 2 to 6). However, this band was also present in the absence of enzyme (lane 1) and is therefore not an RNase H cleavage product. These results were surprising because previously, E478Q RT was capable of producing an RNase H cleavage product on the intact full-length substrate (50). Without any RNase H activity, E478Q RT was unable to proceed with strand transfer. It was therefore possible that E478Q RT was not capable of catalyzing the polymerase-independent RNase H cleavages required for strand transfer in this modified assay.

Complementation of HIV-1 RT and E478Q RT with *E. coli* RNase H. To further characterize the defects of E478Q RT with regard to its RNase H cleavage, it was necessary to determine that this mutated enzyme could perform strand transfer under these modified conditions with the truncated substrates. Therefore, both HIV-1 RT and E478Q RT were complemented with *E. coli* RNase H in a strand transfer reaction. The reactions were performed in the presence of Mg^{2+}, dNTPs, and acceptor; *E. coli* RNase H was added after the 12-min time point for both enzymes. Under these conditions, both the WT HIV-1 RT and *E. coli* RNase H were active and their products could be distinguished. The WT HIV-1 RT products (Fig. 4A, lanes 3, 9, and 15) were similar to those identified in Fig. 2 (lanes 3, 9, and 15). Addition of *E. coli* RNase H after 12 min resulted in much more extensive RNase H degradation (Fig. 4, lanes 4, 10, and 16, indicated by vertical arrows). For E478Q RT, the visible cleavage activity can only be a result of *E. coli* RNase H since the reactions were performed in Mg^{2+} alone. The results of time course analyses using HIV-1 RT and E478Q RT are shown in Fig. 4A and B, respectively. Complementation of HIV-1 RT with *E. coli* RNase H was similar to the results (lanes 1 to 6, lanes 7 to 12, and lanes 13 to 18) with HIV-1 RT alone. A 70-mer strand transfer product was produced in each case. With HIV-1 RT, only a single RNase H cleavage product was observed for the Δ12 construct (Fig. 2). However, addition of *E. coli* RNase H yielded RNA products as small as a diribonucleotide (Fig. 4A, lane 16).

Complementation of E478Q RT with *E. coli* RNase H on the Δ6, Δ9, and Δ12 substrates (Fig. 4B, lanes 1 to 6, lanes 7 to 12, and lanes 13 to 18, respectively) did indeed result in the 70-mer strand transfer product. These results confirmed the requirement of RNase H activity for strand transfer to take place. These experiments also reinforce the finding with HIV-1

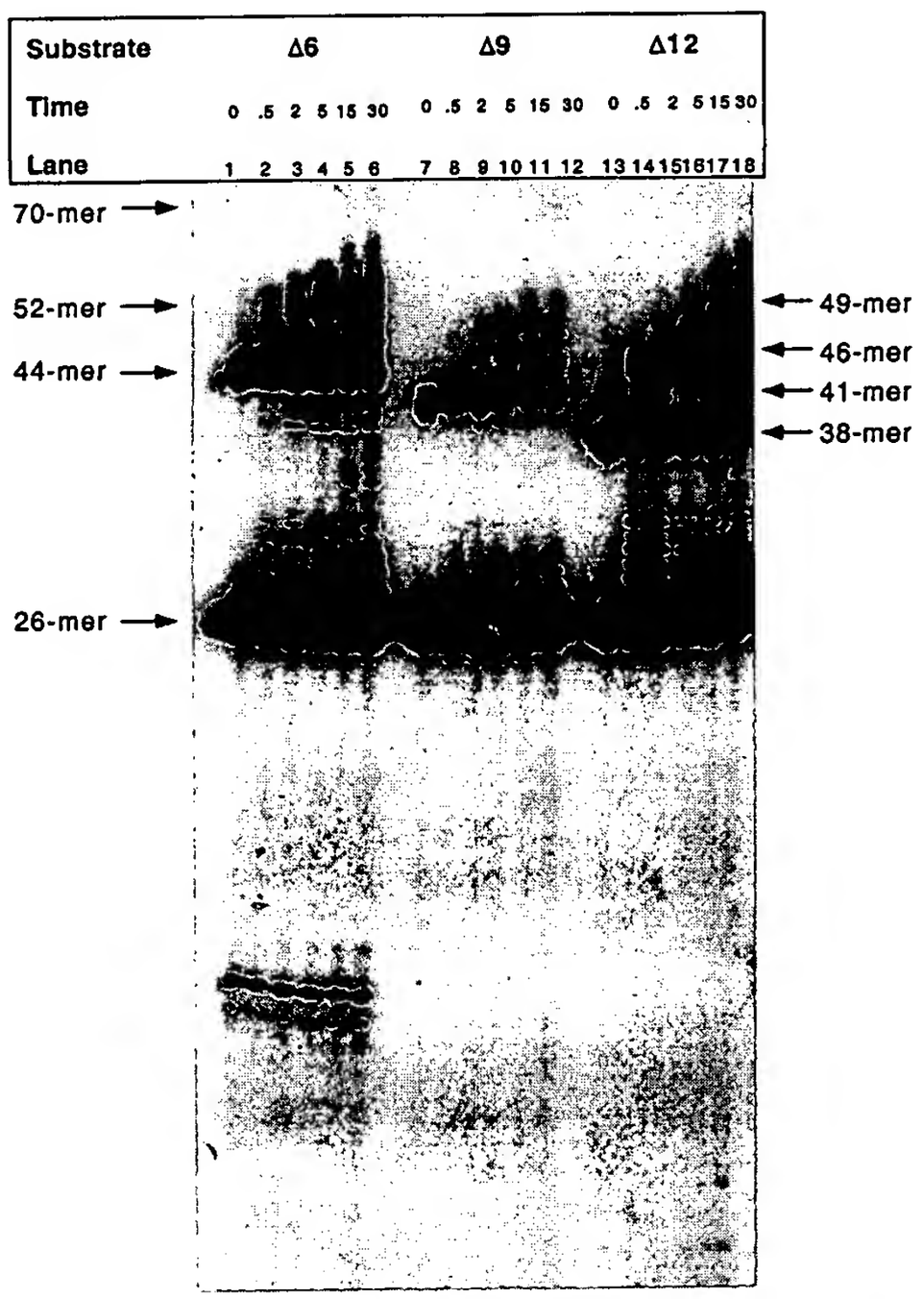


FIG. 3. Truncated substrates assayed with E478Q RT. Reactions were performed as described in Materials and Methods. Lanes 1 to 6, 7 to 12, and 13 to 18 represent E478Q RT assayed with the Δ6, Δ9, and Δ12 constructs, respectively. Time points are indicated above each lane in minutes. Strand transfer products (70-mer), DNA primer (26-mer), and RNase H products are indicated by arrows. Input substrates for the Δ6, Δ9, and Δ12 constructs are 44-mer, 41-mer, and 38-mer, respectively. Initial RNase H cleavage products for the Δ6, Δ9, and Δ12 constructs are 11-mer, 8-mer, and 5-mer, respectively.

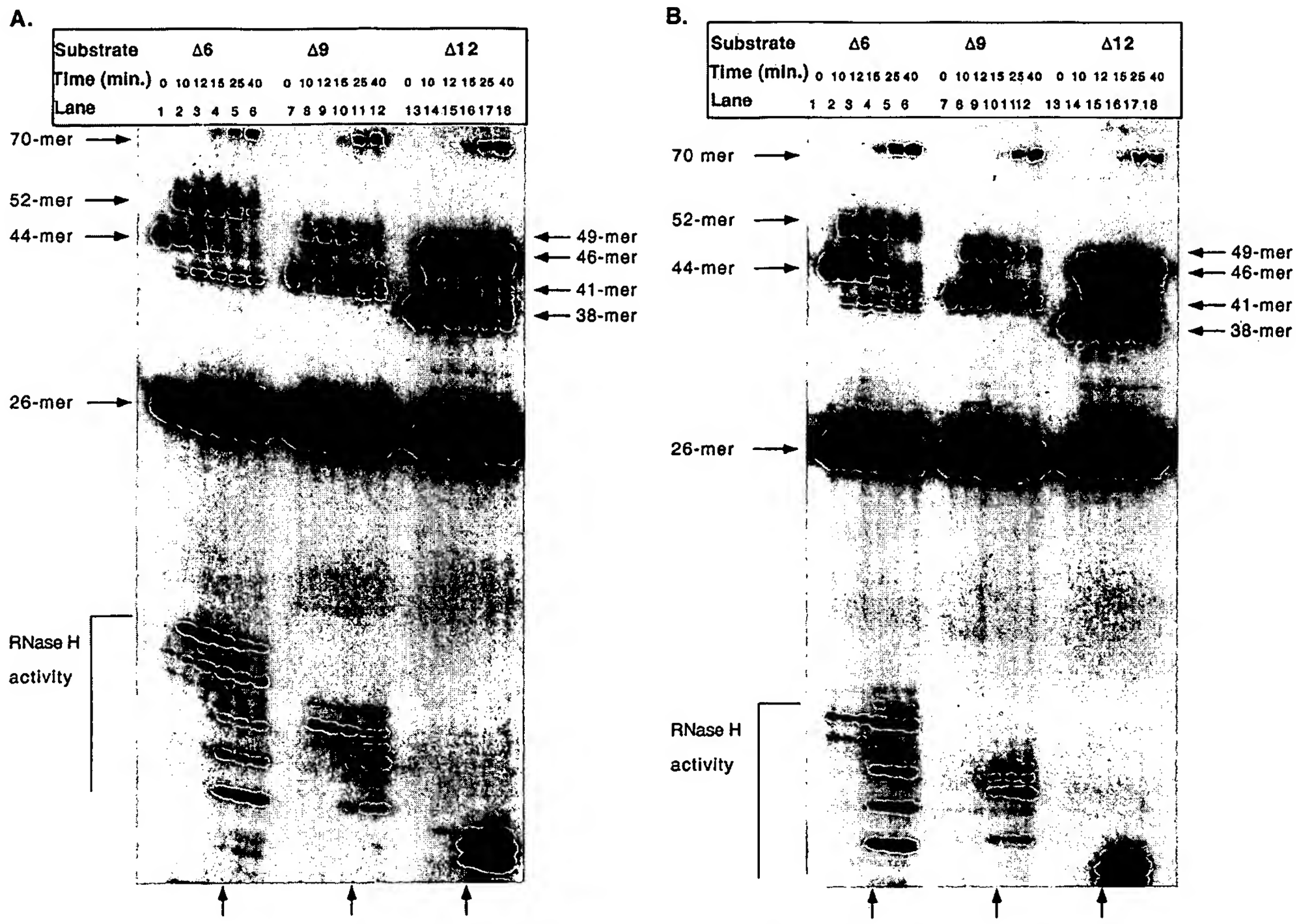


FIG. 4. (A) Complementation of HIV-1 RT with *E. coli* RNase H. Reactions were performed as described in Materials and Methods. Input RNA-DNA, DNA primer, and RNase H cleavage products are indicated by arrows. Reactions were allowed to proceed for 12 min in the presence of Mg^{2+}, and then of *E. coli* RNase H was added (indicated by the vertical arrows). Time points are indicated above each lane in minutes. (B) Complementation of E478Q RT with *E. coli* RNase H. Reactions were performed as described for panel A. Time points are indicated above each lane in minutes. Input RNA-DNA, DNA primer, and RNase H cleavage products are indicated by arrows.

RT that the 5-base overlap between donor and acceptor was sufficient for strand transfer to proceed.

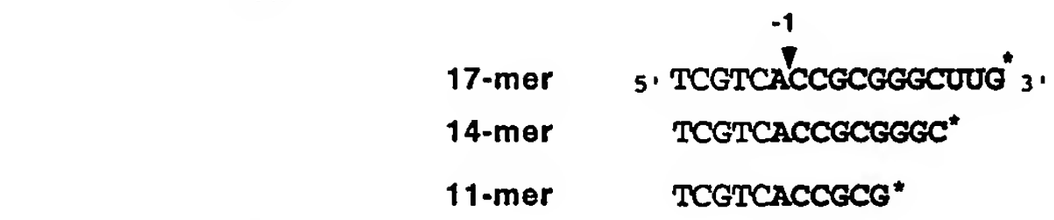
E478Q RT in a polymerase-independent assay. Previous characterization of E478Q indicated that it lacks directional RNase H processing (10) during polymerization. However, the question remained whether E478Q RT can perform any polymerase-independent RNase H cleavages. To test this, a series of short, truncated substrates were generated (Fig. 5A). If RNase H cleavage occurred at the initial (−1) position (indicated by the arrow) (51, 53), substrate binding could not be dictated by the positioning of the 3′OH in the polymerase active site. These RNA substrates contained the same 5′-terminal truncations as the RNAs in the strand transfer assay. Additionally, these substrates truncated the DNA in the RNA-DNA oligonucleotide to 5 nucleotides (Fig. 5A). The RNA-DNA strands were hybridized to DNA oligonucleotides 17, 14, and 11 nucleotides in length, yielding blunt-ended substrates.

Figure 5B illustrates these substrates assayed with HIV-1 RT or E478Q RT. Figure 5B, lanes 1 to 5, 6 to 10, and 11 to 15 represent HIV-1 RT assayed with the truncated substrates. For the 17-mer (lanes 1 to 5), 14-mer (lanes 6 to 10), and 11-mer (lanes 11 to 15) substrates, the correct cleavage products at the −1 position were observed, as indicated by the arrows. These cleavage events have been extensively characterized for HIV-1 RT and an isolated HIV-1 RNase H domain with a related substrate containing a complementary oligonucleotide carrying the entire 18-nucleotide PBS sequence (49). The 17-mer, 14-mer, and 11-mer substrates assayed with E478Q RT are shown in Fig. 5B, lanes 16 to 20, 21 to 25, and 26 to 30, respectively. E478Q RT was capable of cleaving the 17-mer and 14-mer constructs at the predicted −1 position. In contrast to WT RT, E478Q RT did not cleave the 11-mer construct. These results indicate that on defined substrates, E478Q RT is capable of polymerase-independent cleavages.

RNase H cleavage analysis of the truncated DNA substrates. E478Q RT-RNase H recognized and cleaved the blunt 17-mer and 14-mer RNA-DNA hybrid substrates (Fig. 5B) but was inactive on the equivalent 52-mer (Δ6) and 49-mer (Δ9), RNA-DNA hybrid resulting after polymerization (Fig. 3). Two key differences between these substrates are the size of the DNA-DNA hybrid segment and the position of a 3′OH group with respect to the RNA-DNA hybrid. A large DNA substrate may lock the 3′OH within the polymerase active site, whereas a small substrate may have sufficient flexibility to permit binding of the substrate into the RNase H active site.

To address this, a series of substrates were constructed which varied the length of the double-stranded (ds) U5 DNA associated with the tRNA-DNA mimic (Fig. 6A). The sizes of

A.



B.

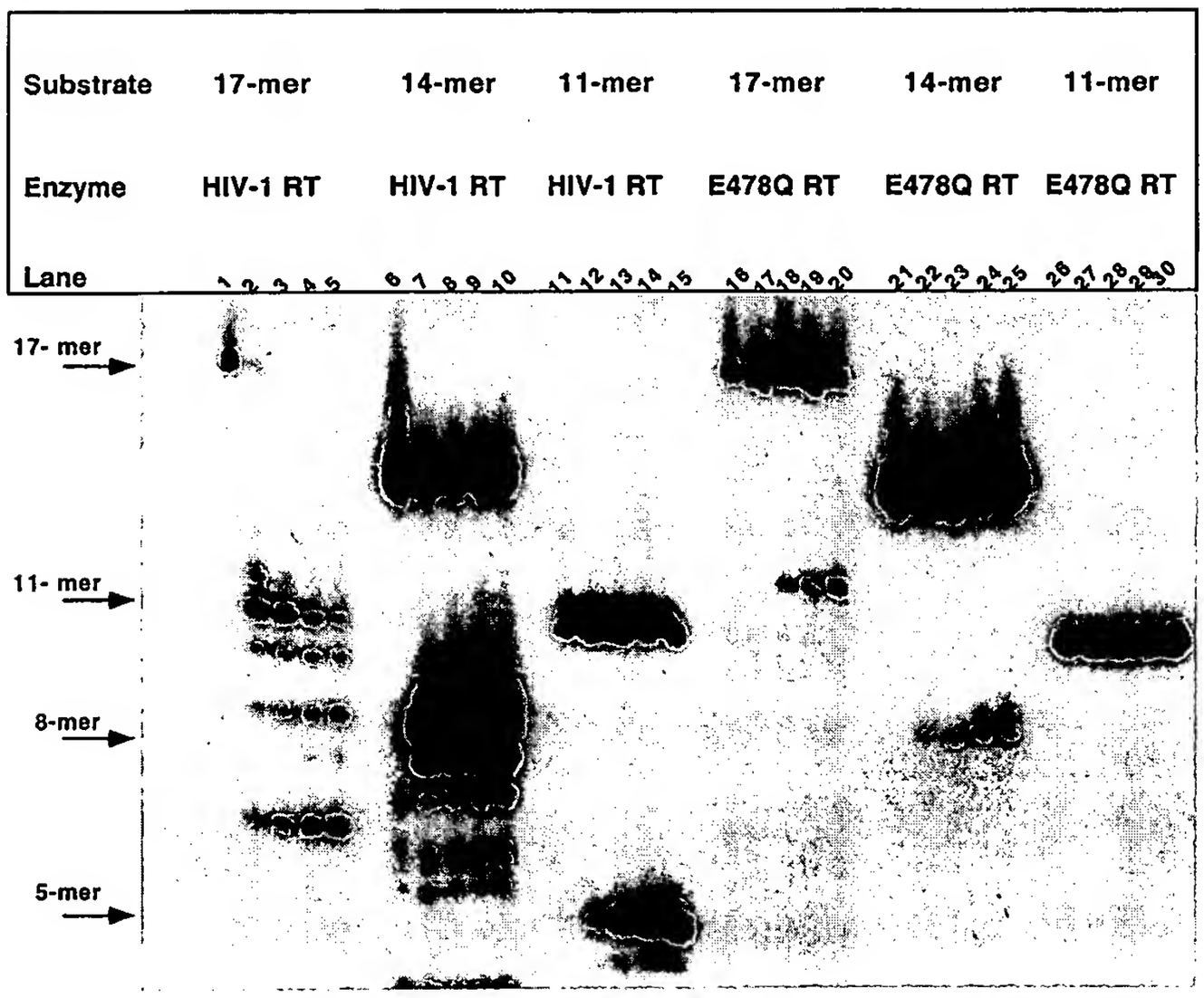


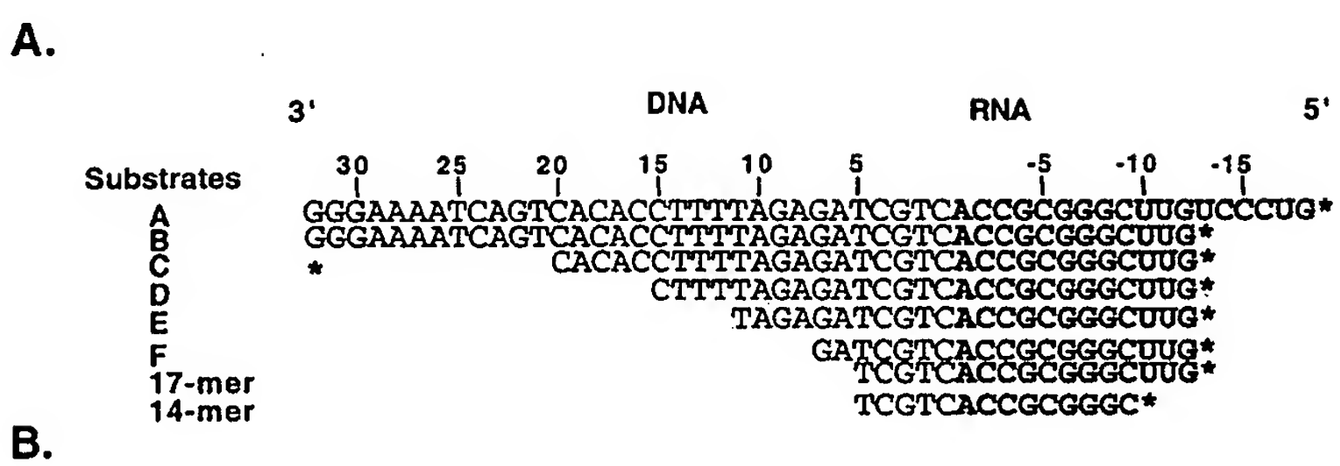
FIG. 5. E478Q RT assayed in polymerase-independent assay. (A) The substrates utilized are illustrated. Substrates were labeled at the 5′ termini with [γ-^{32}P]ATP and prepared as described in Materials and Methods. An asterisk indicates the radiolabel. The RNA portion is indicated in bold. (B) HIV-1 RT assayed with 17-mer (lanes 1 to 5), 14-mer (lanes 6 to 10), 11-mer (lanes 11 to 15); E478Q RT assayed with 17-mer (lanes 16 to 20), 14-mer (lanes 21 to 25), and 11-mer (lanes 26 to 30). Time points are 0, 2, 5, 15, and 30 min for each set of lanes. Reactions were performed as described in Materials and Methods.

the ds DNAs ranged from those that were cleaved, (17-mer; 5-mer DNA plus 12-mer RNA) to those not recognized by E478Q (substrate B; 32-mer DNA plus 12-mer RNA). E478Q RT was capable of recognizing and cleaving a 50-mer RNA-DNA hybrid substrate containing the intact 18-mer RNA portion plus a 32-mer of DNA (substrate A). However, when the RNA portion of the substrate was truncated by 6 nucleotides (substrate B), RNase H cleavage was inhibited. Substrates were constructed in which the DNA-DNA (ds DNA portion) size varied but the size of the RNA-DNA hybrid remained constant (12-mer RNA). Substrates C, D, E, and F possessed 20, 14, 11, and 7 nucleotides, respectively, of ds DNA. The 17-mer (5-mer DNA plus 12-mer RNA) and 14-mer substrates (5-mer DNA plus 9-mer RNA) were shown to be cleaved by E478Q RT (see above) (Fig. 5B).

HIV-1 RT and E478Q RT were assayed for their abilities to cleave these substrates in a polymerase-independent assay (Fig. 6B and C, respectively). For HIV-1 RT, all of the newly constructed substrates (substrates B to F) were cleaved, indicating that the substrates were intact RNA-DNA hybrids (Fig. 6B). Since all of the substrates contained 12 nucleotides of RNA, the RNase H cleavage sites were identical and an 11-mer RNA cleavage product was observed for each substrate. In contrast, with E478Q RT (Fig. 6C), the only substrate that was efficiently cleaved was substrate F, which possessed 7 nucleotides in the DNA-DNA portion and 12 nucleotides in the RNA-DNA hybrid portion of the substrate. The total size of this substrate was 19-mer (7-mer DNA plus 12-mer RNA), which approached that required for polymerase-dependent RNase H cleavages. This was only 2 nucleotides larger than the 17-mer substrate previously cleaved by E478Q RT (Fig. 5B). This result indicates that E478Q RT can perform polymerase-independent RNase H cleavages only on defined small substrates. It is of interest that these substrates are equal to or less than the length defined between the polymerase and RNase H active sites.

DISCUSSION

Two different modes of RNase H activity have been characterized for HIV-1 RT: polymerase dependent and polymerase independent (20). Polymerase-dependent RNase H cleavages are those that occur while the polymerase domain is actively synthesizing. The RNase H active site lags approximately 18 to 20 nucleotides behind the polymerase active site, which has been characterized through footprinting analysis for HIV-1 RT and Moloney murine leukemia virus RT (55, 57). On the plus-strand strong stop tRNA-DNA, tRNA removal occurs while the RT is paused trying to use the modified tRNA residue as template. Although the polymerase is not actively synthesizing, the initial RNase H cleavage can be viewed as polymerase dependent since the nucleic acid substrate remains bound in



B.

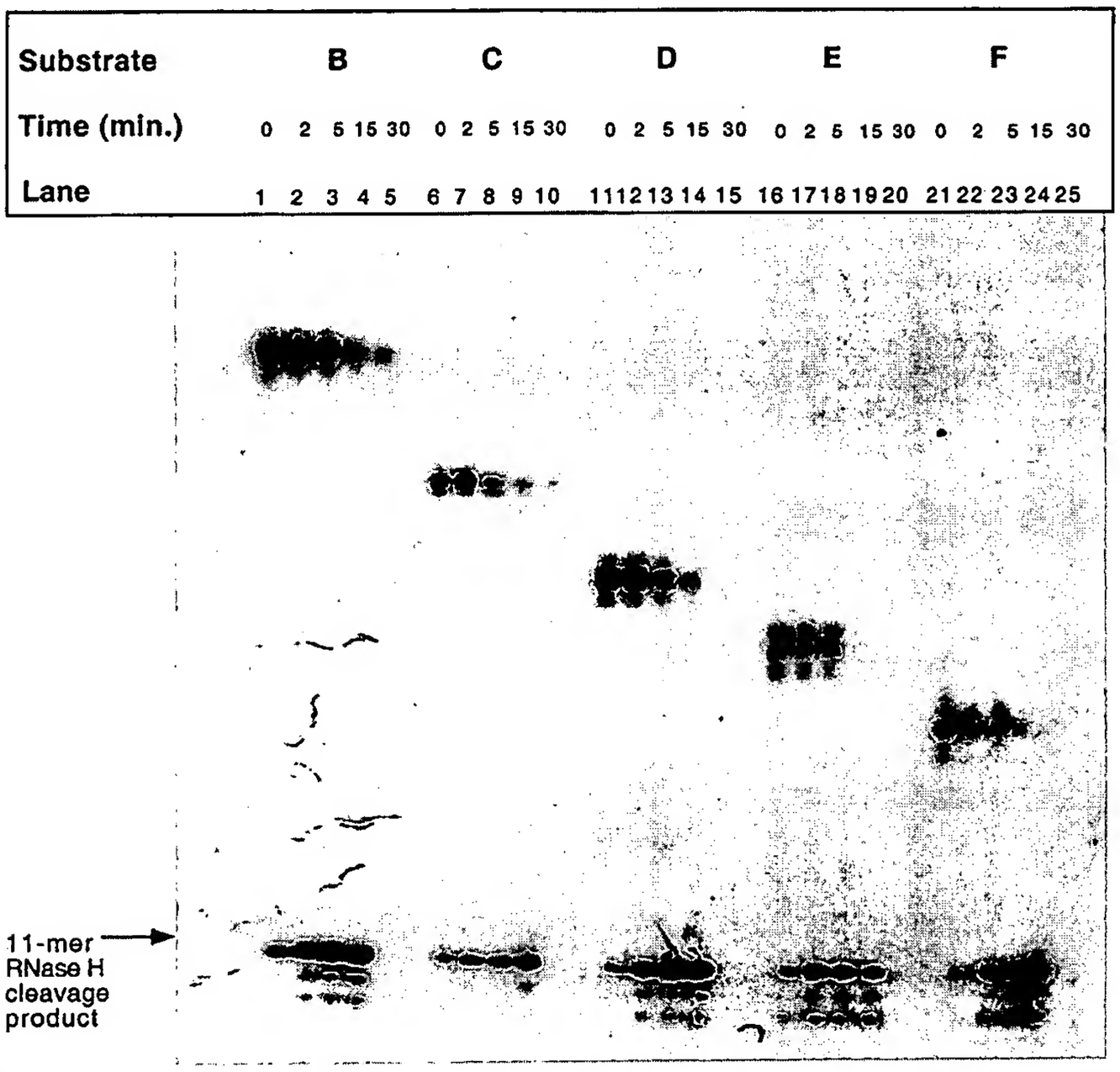


FIG. 6. RNase H cleavage analysis of truncated DNA substrates. (A) The substrates utilized are illustrated and are labeled A through F. The RNA portions are indicated in bold, and an asterisk indicates the radiolabel. The substrates were prepared as described in Materials and Methods. (B) Substrates B to F assayed with HIV-1 RT. Reactions were performed as described in Materials and Methods. Time course reactions are shown, and time points are indicated above each lane, along with the substrate utilized. The RNase H cleavage product is designated and indicated by an arrow. (C) Substrates B to F assayed with E478Q RT. Reactions were performed as described in Materials and Methods. Time course reactions are shown, and time points are indicated above each lane, along with the substrate utilized. The RNase H cleavage product is designated and indicated by an arrow.

the polymerase active site. The initial RNase H cleavage occurs at the penultimate nucleotide of the tRNA (53). Polymerase-independent cleavages are less well understood and may represent different types of cleavages. These cleavages result in the production of smaller products (22) and have been defined as "directional processing of the RNA primer" (10). Although secondary to the initial cleavage during reverse transcription, these events are required for the ultimate release of the RNA primer. Mutants with defects in the ability to perform this function are unable to perform strand transfer (10, 21, 23, 50). Polymerase-independent RNase H cleavages require a second binding event by RT (23). This "rebinding" event may require a change between the type of conformation required for active DNA synthesis and polymerase-dependent RNase H activity.

We have developed an in vitro assay which requires polymerase-independent RNase H activity during the second-strand transfer reaction. Previously, we utilized substrates which mimic the U5/PBS border, resulting from plus-strand strong stop DNA synthesis. The original substrates possessed either an 18-mer RNA oligonucleotide or the intact $tRNA^{Lys,3}$ (50). With those substrates, the RNase H activity we observed was most probably a combination of polymerase-dependent and -independent cleavages. The substrates utilized in this study have truncations in the 5′ RNA portion and have a maximum length of 12 ribonucleotides. Therefore, cleavage of these substrates would require polymerase-independent RNase H activity, due to the suboptimal distance between the polymerase and RNase H active sites.

C.

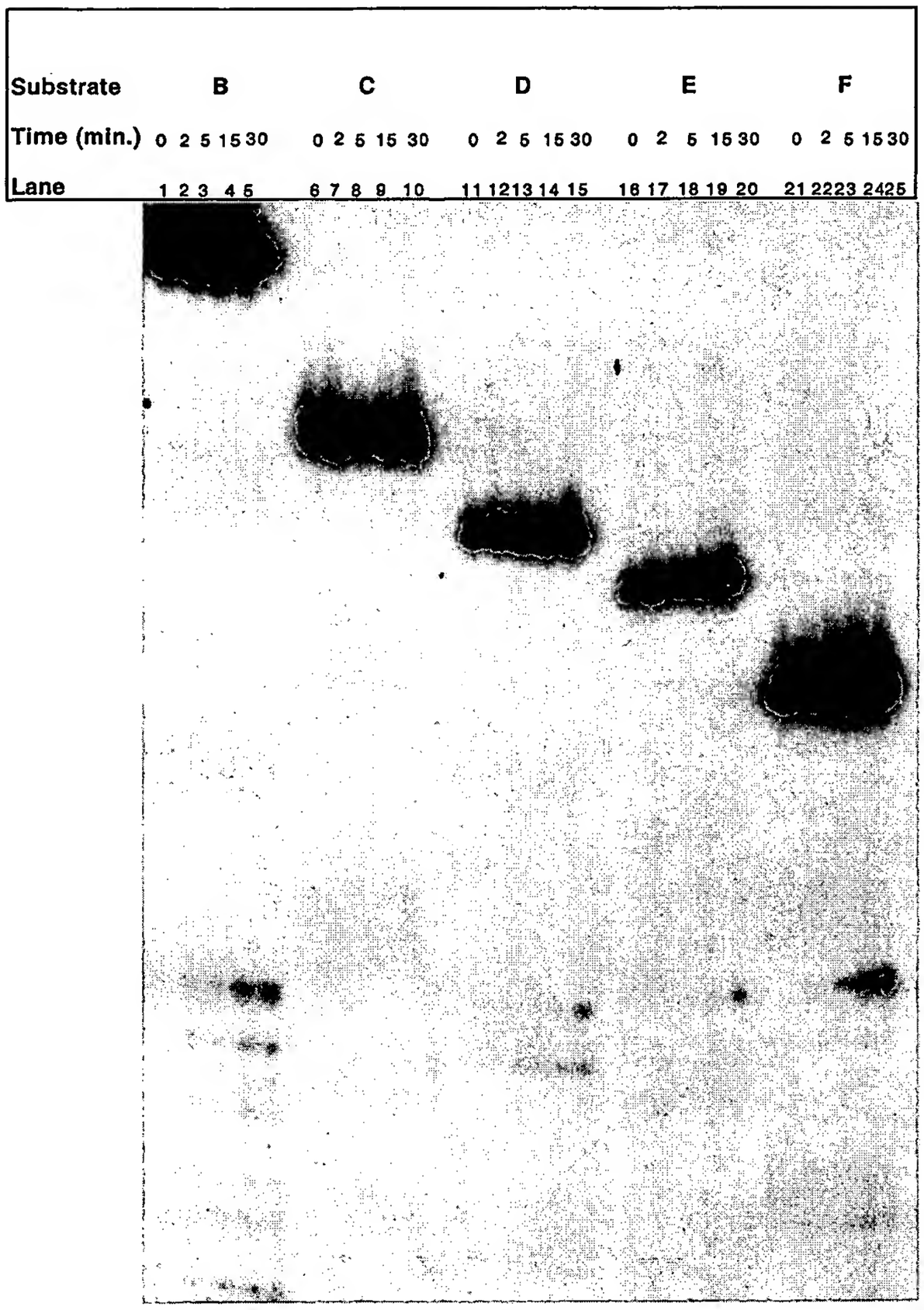
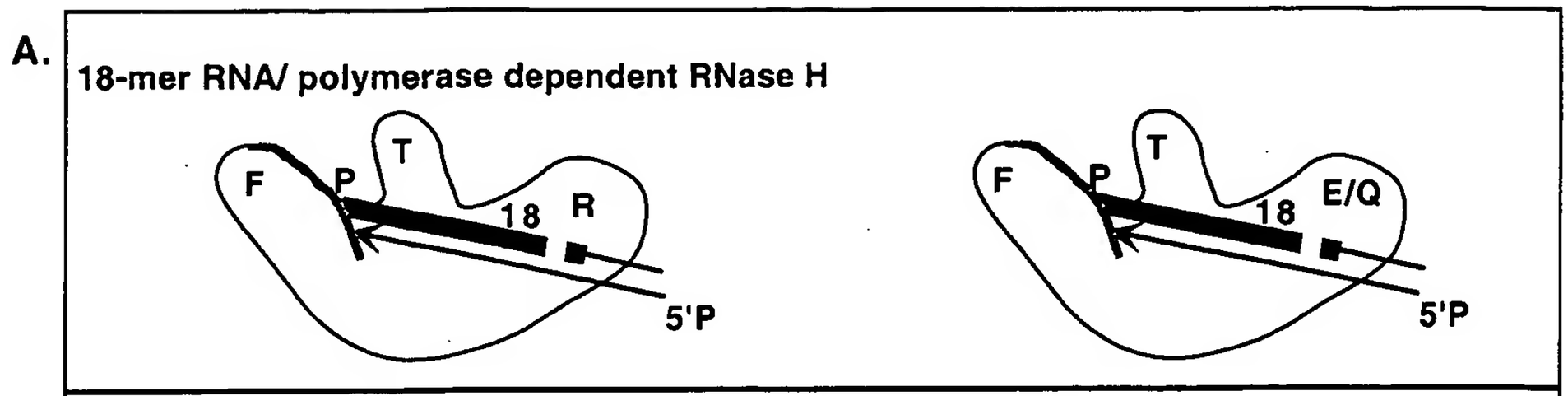
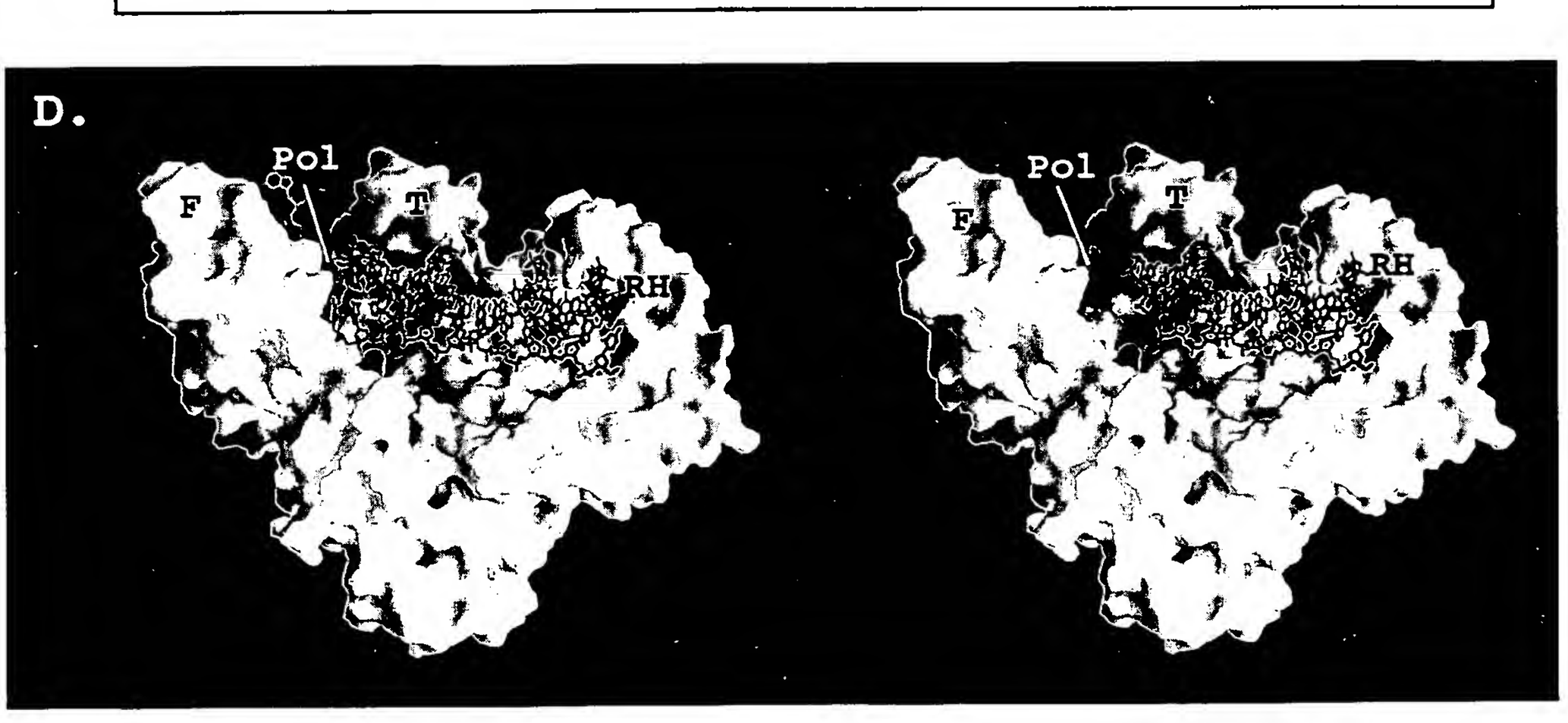


FIG. 6—*Continued*.

A model illustrating the possible binding conformations adopted by WT RT and E478Q RT is shown in Fig. 7. This model is based on the crystal structure of a covalently trapped catalytic complex of RT with DNA (24) (Fig. 7D, left panel). In this complex, the ds DNA spans 23 nucleotides and approximates the 12-mer RNA plus 11-mer DNA substrates used in the present study. In Fig. 7D, left panel, the substrate is positioned in the polymerase active site (labeled Pol). In Fig. 7D, right panel, the substrate was modified such that a potential RNA strand (red) is within the RNase H active site (RH) and extended 12 nucleotides toward the polymerase active site. This substrate lacks the stabilizing contacts within the thumb domain. The extension of the substrate exiting the RNase H domain was not modeled due to limited structural data on these contacts. Figure 7A to C utilize a schematic of this molecular model to summarize the results presented in this study. On a full-length model substrate (50-mer: 32-mer DNA plus 18-mer RNA) (Fig. 7A), both E478Q RT and WT HIV-1 RT bind in a polymerase-dependent conformation and produce the well-characterized −1 RNase H cleavage product (depicted as a nick within the RNA [thick bar]) (50). The size of the RNA portion of the substrate allows correct positioning of the RNase H active site near the RNA-DNA junction. This RNase H cleavage event is driven by the polymerase domain and is guided by the 5′ phosphate of the RNA and the 3′ OH of the DNA (39).

Figure 7B represents the truncated substrate (44-mer: 32-mer DNA plus 12-mer RNA). This substrate is suboptimal for a polymerase-dependent mode of binding due to its truncation in the 5′ portion of the RNA. Despite this, the WT RT maintains its RNase H cleavages on these substrates. This implies that it is capable of the release and rebinding of these substrates (23). E478Q RT, which is compromised in the RNase H domain, has regained the single endoribonucleolytic cleavage ability only in the presence of Mn^{2+} (10). E478Q RT binds in the same polymerase-dependent orientation; however, RNase H cleavage never occurs because the RNase H active site is never positioned near the RNA-DNA junction on this sub-


A.
18-mer RNA/ polymerase dependent RNase H
F
P
T
18
R
5'P
F
P
T
18
E/Q
5'P
B.
12-mer RNA/ polymerase dependent RNase H
F
P
T
12
R
5'P
F
P
T
12
R
5'P
F
P
T
12
E/Q
5'P
F
P
T
12
E/Q
5'P
C.
12-mer RNA/ polymerase independent RNase H
F
P
T
12
E/Q
11-32
5'P
F
P
T
12
E/Q
5-7
5'P


D.
Pol
F
T
RH
Pol
F
T
RH

strate. It appears that E478Q in the presence of Mn^{2+} has not maintained the ability to communicate the release and subsequent rebinding to these large substrates, which possess a larger portion of ds DNA than the smaller substrates do. Alternatively, catalysis for the polymerase-independent cleavages may be slower than dissociation on these substrates (12). In the polymerase-dependent reaction catalyzed by E478Q, the RNA-DNA hybrid remains accessible to further degradation by exogenously added *E. coli* RNase H. This indirectly implies that the enzyme is released from the substrate. Complementation with *E. coli* RNase H is sufficient for subsequent strand transfer catalyzed by E478Q RT.

When presented with a substrate truncated in the DNA portion as well as in the RNA portion (19-mer: 7-mer DNA plus 12-mer RNA), E478Q RT was capable of performing an RNase H cleavage. This substrate (Fig. 7C) is small enough that it is not locked into the polymerase active site. Therefore, a polymerase-independent or RNase H active site-driven cleavage may occur. HIV-1 RT preferentially binds a heteroduplex RNA-DNA region with a greater affinity than it binds ds DNA (12), which may be directing the RNase H cleavage. However, once the DNA portion was increased to 11 nucleotides (23-mer: 11-mer DNA plus 12-mer RNA), the substrate was again locked into the polymerase active site and no RNase H activity was observed. This indicates that the overall substrate size must be smaller than the distance between the polymerase and RNase H active sites (55, 57) in order for E478Q RT to perform these smaller polymerase-independent RNase H cleavages. The E478Q mutation may alter the ability of the enzyme to translocate across larger substrates to position itself correctly for RNase H cleavage, whereas a smaller substrate can slide into the RNase H active site.

Several factors have been identified which contribute to the positioning of the template. These include the position of the primer 3′OH, the position of the 5′ phosphate of the RNA, and the minor groove binding tract within the thumb subdomain. We propose that with substrates longer than 19 nucleotides, a polymerase-dependent conformation dominates, whereas with the smaller substrates, binding to either the polymerase or RNase H active sites may occur. This may be a combination of the high affinity of the polymerase domain for the termini of the substrate and a result of the defect in the RNase H domain of E478Q RT. With E478Q, subsequent rebinding to the substrate may remain driven solely by the recognition of the termini of the substrate. Rebinding (23) would occur in the same location and would never position the RNA-DNA substrate in a conformation required for a polymerase-independent cleavage. It is therefore surprising that the smaller substrates are capable of binding E478Q within the RNase H domain. This implies that there may be a stabilizing effect of having a substrate exit the RNase H domain. Reports have suggested that substrate binding by the RNase H domain contributes to processive DNA synthesis (8).

Studies have been performed which show that the 5′ end of the RNA sequence can be responsible for directing its cleavage by HIV-1 RNase H (39, 40). Those studies suggested that RT is capable of binding the substrate in two different ways: either based on the 3′ end of the DNA strand or based on the 5′ end of the RNA strand. The binding was dependent on the substrate structure and on whether the DNA or RNA ends were recessed. When binding was driven by the position of the RNA 5′ phosphate, RNase H cleavage still occurred 18 nucleotides upstream, maintaining the optimal spatial distance between the two active sites. However, with the substrates utilized in our study, binding of the 5′ phosphate within the polymerase active site would position the RNase H active site within the ds DNA region. WT RT is able to overcome this obstacle; therefore, this cannot be the dominant feature, since cleavage at the penultimate nucleotide within the RNA was consistently observed.

Another contributing factor may be the role of the p66 thumb subdomain in positioning the RNA-DNA substrate. In the crystal structure of HIV-1 RT bound with nonnucleoside inhibitor, the position of the thumb differs from that of the inhibitor free enzyme (11, 15, 16, 45). The α-helix H within the thumb acts as a minor groove binding tract (5, 6). This helix plays a role in the binding of primer-template complexes. The RT mutant W266A (mutated within α-helix H) lost the ability to position the 3′OH/5′ phosphate within the polymerase active site and resulted in imprecise removal of the polypurine tract primer (43). Molecular modeling of a truncated substrate in the RNase H active site precluded the interaction of a 12-base extension with the minor groove binding tract (Fig. 7D, right). Flexibility of the thumb and/or perturbation of the size of the major and minor groove of the substrate could alter these interactions. Cooperativity of the thumb with either the polymerase or RNase H active sites could therefore be a discriminating factor for the positioning of the tRNA-DNA substrate.

We propose two different binding conformations for HIV-1 RT in removing the tRNA, a polymerase-dependent binding mode and a polymerase-independent binding mode. Differential conformations for endonuclease and exonuclease RNase H activities were previously described (60). Another study demonstrated that in the switch from initiation to elongation, the RT must dissociate before efficient and processive elongation can occur (31). That study also concluded that the length of the synthesized DNA affects this switch, since it would change the proximity of the RNA/DNA junction on the primer strand relative to α-helix H. The truncated substrates utilized in our present study would place the RNase H active site within the U5 DNA portion once synthesis is complete; therefore, dissociation would have to occur for any RNase H cleavage to occur. HIV-1 RT has no defects in its ability to perform this switch and to cleave truncated substrate intermediates. However, the E478Q RT was unable to carry out these cleavages on substrates that were identical in size.

The present study has also determined that an overlap of 5

FIG. 7. Model of WT RT and E478Q RT polymerase-dependent and polymerase-independent RNase H activities. The possible binding orientations for HIV-1 RT and E478Q RT are shown. (A to C) WT RT and E478Q RT are shown positioned on the various substrates, 50-mer substrate (32-mer DNA, 18-mer RNA) (A), 44-mer substrate (32-mer DNA, 12-mer RNA) (B), and 19-mer (7-mer DNA, 12-mer RNA) (C). The RNA portion is indicated by the thick line, and the DNA is indicated by the solid black line. The position of RNase H cleavage is indicated by a nick in the RNA strand. The WT and E478Q RT can be distinguished by the presence of an R (WT) or E/Q (E478Q RT) in the RNase H domain. Additionally, the thumb and polymerase domains are indicated by T and P, respectively. The 5′ phosphate is indicated, as well as the size of the RNA on each model substrate. (D) Models of substrates bound in the polymerase active site (left) and the RNase H active site (right) in HIV-1 RT. The electrostatic potential mapped on the molecular surface rendering of the HIV-1 RT (GRASP [38]) is shown with the template-primer as bound in the structure reported by Huang et al. (24) (1rtd), and substrates are shown as stick models. Positively charged amino acids are shown in blue, and negatively charged amino acids are shown in red. In the right-hand panel, the substrate found in the 1rtd structure has been truncated to include only 12 bp of template-primer extending from the RNase H active site. This truncated substrate makes very limited interactions with the thumb. T, thumb; F, fingers; RH, RNase H active site; Pol, polymerase active site.

mann, and R. Marquet. 1999. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. EMBO J. 18:1038–1048.
28. Jiang, M., J. Mak, A. Lahda, E. Cohen, M. Klein, B. Rovinski, and L. Kleiman. 1993. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. J. Virol. 67:3246–3253.
29. Jones, J. S., R. A. Allan, and H. M. Temin. 1994. One viral RNA is sufficient for synthesis of viral DNA. J. Virol. 68:207–216.
30. Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice, and T. A. Steitz. 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256:1783–1790.
31. Lanchy, J., G. Keith, S. F. J. Le Grice, B. Ehresmann, C. Ehresmann, and R. Marquet. 1998. Contacts between reverse transcriptase and the primer strand govern the transition from initiation to elongation of HIV-1 reverse transcription. J. Biol. Chem. 273:24425–24432.
32. Le Grice, S. F. J., and F. Gruninger-Leitch. 1990. Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography. Eur. J. Biol. 187:307.
33. Li, X., J. Mak, E. J. Arts, Z. Gu, L. Kleiman, M. A. Wainberg, and M. A. Parniak. 1994. Effects of alterations of primer-binding site sequences on human immunodeficiency virus type 1 replication. J. Virol. 68:6198–6206.
34. Lightfoot, M. M., J. E. Coligan, T. M. Folks, A. S. Fauci, M. A. Martin, and S. Venkatesan. 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of acquired immunodeficiency syndrome virus. J. Virol. 60:771–775.
35. Mak, J., M. Jiang, M. A. Wainberg, M.-L. Hammarskjold, D. Rekosh, and I. Kleiman. 1994. Role of Pr160$^{gag\text{-}pol}$ in mediating the selective incorporation of tRNA$^{(Lys)}$ into human immunodeficiency virus type 1 particles. J. Virol. 68:2065–2072.
36. Mak, J., A. Khorchid, Q. Cao, Y. Huang, I. Lowy, M. A. Parniak, V. R. Prasad, M. A. Wainberg, and L. Kleiman. 1997. Effects of mutations in Pr160 gag-pol upon tRNA$^{(Lys3)}$ and Pr160 gag-pol incorporation into HIV-1. J. Mol. Biol. 265:4555–4564.
37. Mizrahi, V., G. M. Lazarus, L. M. Miles, C. A. Meyers, and C. DeBouck. 1989. Recombinant HIV-1 reverse transcriptase: purification, primary structure, and polymerase/ribonuclease H activities. Arch. Biochem. Biophys. 273:347–358.
38. Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296.
39. Palaniappan, C., G. M. Fuentes, L. Rodriguez-Rodriguez, P. J. Fay, and R. A. Bambara. 1996. Helix structure and ends of RNA/DNA hybrids direct the cleavage specificity of HIV-1 reverse transcriptase RNase H. J. Biol. Chem. 271:2063–2070.
40. Palaniappan, C., J. K. Kim, M. Wisniewski, P. J. Fay, and R. A. Bambara. 1998. Control of initiation of viral plus strand DNA synthesis by HIV reverse transcriptase. J. Biol. Chem. 273:3808–3816.
41. Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B. D. Preston, and J. P. Dougherty. 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. J. Virol. 69:7991–8000.
42. Peliska, J. A., and S. J. Benkovic. 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science 258:1112–1118.
43. Powell, M. D., W. A. Beard, K. Bebenek, K. J. Howard, S. F. J. Le Grice, T. A. Darden, T. A. Kunkel, S. H. Wilson, and J. G. Levin. 1999. Residues in the αH and αI helices of the HIV-1 reverse transcriptase thumb subdomain required for the specificity of RNase H-catalyzed removal of the polypurine tract primer. J. Biol. Chem. 274:19885–19893.
44. Powell, M. D., M. Ghosh, P. S. Jacques, K. J. Howard, S. F. L. Grice, and J. G. Levin. 1997. Alanine-scanning mutations in the "primer grip" of p66 HIV-1 reverse transcriptase result in selective loss of RNA priming activity. J. Biol. Chem. 272:13262–13269.
45. Ren, J., R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kiraby, J. Keeling, G. Darby, Y. Jones, D. Stuart, and D. Stammers. 1995. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nat. Struct. Biol. 2:293–308.
46. Roth, M. J., P. Schwartzberg, and S. P. Goff. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence of IN function and terminal DNA sequence. Cell 58:47–54.
47. Schatz, O., F. V. Cromme, F. Gruninger-Leitch, and S. F. J. Le Grice. 1989. Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. FEBS Lett. 237:311–314.
48. Smith, C. M., W. B. Potts III, J. S. Smith, and M. J. Roth. 1997. RNase H cleavage of tRNA^{Pro} mediated by M-MuLV and HIV-1 reverse transcriptases. Virology 229:437–446.
49. Smith, C. M., O. Leon, J. S. Smith, and M. J. Roth. 1998. Sequence requirements for removal of tRNA by an isolated HIV-1 RNase H domain. J. Virol. 72:6805–6812.
50. Smith, C. M., J. S. Smith, and M. J. Roth. 1999. RNase H requirements for the second strand transfer reaction of HIV-1 reverse transcription. J. Virol. 73:6573–6581.
51. Smith, J. S., S. Kim, and M. J. Roth. 1990. Analysis of long terminal repeat circle junctions of human immunodeficiency virus type 1. J. Virol. 64:6286–6290.
52. Smith, J. S., and M. J. Roth. 1993. Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. J. Virol. 67:4037–4049.
53. Smith, J. S., and M. J. Roth. 1992. Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA^{Lys3}. J. Biol. Chem. 267:15071–15079.
54. Wakefield, J. K., and C. D. Morrow. 1996. Mutations within the primer binding site of the human immunodeficiency virus type 1 define sequence requirements essential for reverse transcription. Virology 220:290–298.
55. Wohrl, B. M., B. Ehresmann, G. Keith, and S. F. Le Grice. 1993. Nuclease footprinting of human immunodeficiency virus/tRNA$^{(Lys\text{-}3)}$ complex. J. Biol. Chem. 268:13617–13624.
56. Wohrl, B. M., M. M. Georgiadis, A. Telesnitsky, W. A. Hendrickson, and S. F. Le Grice. 1995. Footprint analysis of replicating murine leukemia virus reverse transcriptase. Science 267:96–99.
57. Wohrl, B. M., C. Tantillo, E. Arnold, and S. F. Le Grice. 1995. An expanded model of replicating human immunodeficiency virus reverse transcriptase. Biochemistry 34:5343–5356.
58. Wu, T., J. Guo, J. Bess, L. E. Henderson, and J. G. Levin. 1999. Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. J. Virol. 73:4794–4805.
59. Zhan, X., and R. J. Crouch. 1997. The isolated RNase H domain of murine leukemia virus reverse transcriptase. Retention of activity with concomitant loss of specificity. J. Biol. Chem. 272:22023–22029.
60. Zhan, Z., C.-K. Tan, W. A. Scott, A. M. Mian, K. M. Downey, and A. G. So. 1994. Catalytically distinct conformations of the ribonuclease H of HIV-1 reverse transcriptase by substrate cleavage patterns and inhibition by azidothymidylate and N-ethylmaleimide. Biochemistry 33:1366–1372.

Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase

Paul L. Boyer*†, Stefan G. Sarafianos‡, Edward Arnold‡, and Stephen H. Hughes*†§

*Advanced BioScience Laboratories-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, MD 21702-1201; and ‡Center for Advanced Biotechnology and Medicine and Rutgers University Chemistry Department, 679 Hoes Lane, Piscataway, NJ 08854-5638

Communicated by George F. Vande Woude, Van Andel Research Institute, Grand Rapids, MI, December 31, 1999 (received for review April 25, 1999)

We have examined amino acid substitutions at residues 115 and 116 in the reverse transcriptase (RT) of HIV-1. A number of properties were examined, including polymerization and processivity on both DNA and RNA templates, strand displacement, ribonucleotide misincorporation, and resistance to nucleoside analogs. The RT variants Tyr-115–Phe and Phe-116–Tyr are similar to wild-type HIV-1 RT in most, but not all, respects. In contrast, the RT variant Tyr-115–Val is significantly impaired in polymerase activity compared with wild-type RT; however, Tyr-115–Val is able to incorporate ribonucleotides as well as deoxyribonucleotides during polymerization and is resistant to a variety of nucleoside analogs.

The reverse transcriptase (RT) of HIV-1 is essential for the replication of the virus (for a review, see ref. 1) and is an important target for antiviral therapy. Two major classes of RT inhibitors, nucleoside analogs and nonnucleoside inhibitors, have been discovered. However, mutations within RT engender resistance to both classes of inhibitors (for reviews, see refs. 2–4). To understand in greater detail how HIV-1 RT is able to mutate to render nucleoside analogs less effective, the nucleotide binding site at the polymerase active site must be studied in greater detail.

Based on the available crystal structures of Moloney murine leukemia virus (MMLV) RT and HIV-1 RT, models have been developed for nucleotide binding sites of these two related enzymes. In the available models (5), the amino acids Phe-155 (MMLV RT) and Tyr-115 (HIV-1 RT) interact with the deoxyribose of the incoming dNTP. The structure of a ternary complex of HIV-1 RT, a double-strand DNA, and a bound dTTP now has been determined (6) and shows that, as predicted (5), Tyr-115 interacts directly with the deoxyribose of the incoming triphosphate.

In analyzing mutations at Tyr-115, several groups have found that the only amino acid substitution that has little deleterious effect on the enzyme is Tyr-115–Phe (7–14). Substitution of Tyr-115 with polar amino acids is extremely deleterious (9–13) whereas substitution with hydrophobic amino acids usually has an intermediate effect (7, 8, 11, 12, 14). Recently, the Tyr-115–Phe mutation, along with the mutations Lys-65–Arg, Leu-74–Val, and Met-184–Val, has been detected in HIV-1 RTs resistant to the carbocyclic nucleoside analog 1592U89 (abacavir) (15, 16). The effects of amino acid substitutions at residue 116 have been less well characterized (7, 8), although the amino acid substitution Phe-116–Tyr is one of the mutations present in an RT variant resistant to multiple dideoxynucleotides (17–19). In this study, we have analyzed the effects of mutations Tyr-115–Phe, Tyr-115–Val, and Phe-116–Tyr in the nucleotide-binding pocket on the activity of HIV-1 RT.

Some of the substitutions we have chosen to make in HIV-1 RT represent amino acids found at the equivalent positions in other viral polymerases. As has already been mentioned, there is a phenylalanine in MMLV RT at the position equivalent to position 115 of HIV-1 RT (5), and a tyrosine in hepatitis B RT at the position equivalent to position 116 of HIV-1 RT (20). We also wanted to compare the effects of introducing valine at position 115 of HIV-1 RT with the MMLV RT mutant with valine at position 155 (21, 22).

Materials and Methods

Preparation of p66/p51 Heterodimers. *Bsp*MI cassette mutagenesis was used to introduce mutations into the p66 coding region (7, 23). The DNA segments encoding the "His-tagged" RT variants were cloned into the vector S-D prot, and the clones were then introduced into the *Escherichia coli* strain BL21 (DE3) pLysE (24). The resulting heterodimers were purified by metal chelation chromatography (24).

Polymerase and Inhibitor Assays. Wild-type RT and the variant RTs were analyzed for polymerase activity by using a variety of different template-primer substrates (Fig. 1). As described (26), we constructed a series of plasmids containing the polypurine tract (PPT, 5′-AAAAGAAAAGGGGGGGGA-3′), U3, R, U5, and the primer binding site (PBS, 5′-GTCCCTGTTCGGGCGCCA-3′) from the proviral clone pNL 4–3 (25) in the vectors Litmus 28 and Litmus 29 (Not).

For each enzyme sample to be assayed, 0.5 μg of single-strand DNA and 0.1 μl of 10 OD_{260}/ml oligonucleotide were hybridized. The mixture was adjusted to contain 25 mM Tris·Cl (pH 8.0), 75 mM KCl, 8.0 mM $MgCl_2$, 2 mM DTT, 1 μg/ml acetylated BSA, 10 mM CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), 10 μM each of dATP and dTTP, 5 μM each of dGTP and dCTP, and 0.5 μl each of [α^{32}P]dGTP and [α^{32}P]dCTP in a final volume of 100 μl. The reaction was initiated by the addition of 0.1 μg of enzyme and was allowed to proceed for 1 h. The reactions were terminated by the addition of 10 μl of 10 mg/ml sheared, denatured salmon sperm DNA, and 3 ml of ice-cold 100% trichloroacetic acid (TCA). The precipitate was collected on Whatman GF/C glass filters and counted.

The RNA-dependent DNA-dependent polymerase (RDDP) assays used four template-primer combinations: (*i*) poly(rC)·oligo(dG) obtained from Amersham Pharmacia; (*ii*) poly(rA)·oligo(dT) obtained from Amersham Pharmacia; (*iii*) PPT-PBS antisense RNA hybridized to the PPT DNA oligonucleotide; and (*iv*) PPT-PBS sense RNA hybridized to the PBS DNA oligonucleotide. The last two RDDP assays are similar to the assays described previously (23) except that RNase-free reagents and 1 unit/μl RNase inhibitor (Ambion) were used.

Antisense RNA was generated from the *Spe*I-linearized plasmid PPT-PBS(T_n) by using the T7 MEGAscript kit and protocol (Ambion) as described (26); sense RNA was generated by using the *Not*I-linearized plasmid PPT-PBS(A_n). The full-length RNAs generated from these plasmids have poly(A) tails and were purified by using the PolyAT tract system (Promega).

Abbreviations: RT, reverse transcriptase; MMLV, Moloney murine leukemia virus; DDDP, DNA-dependent DNA polymerase; RDDP, RNA-dependent DNA polymerase; PPT, polypurine tract; PBS, primer binding site.

†Present address: HIV Drug Resistance Program, P.O. Box B, Building 539, Room 130A, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201.

§To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

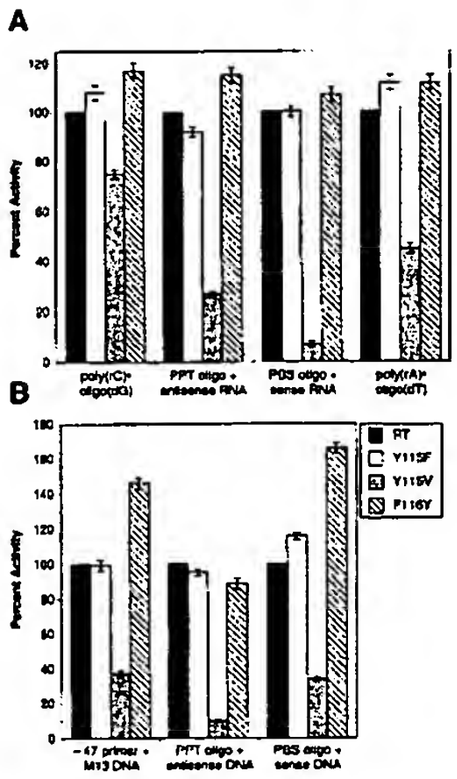


Fig. 1. Polymerase activity of the RT variants relative to wild-type RT. The level of radioactivity incorporated by wild-type RT represents 100% activity and the level of radioactivity incorporated by the RT variants is normalized to this value. The template primers used are described in *Materials and Methods*. (*A*) RDDP activities. (*B*) DDDP activities.

The inhibition of wild-type RT and the RT variants by the nucleoside analogs ddITP and ddGTP was assayed by using poly(rC)·oligo(dG) as the template-primer whereas the nucleoside analogs 3′-azido-3′-deoxythymidine 5′triphosphate (AZTTP) (Moravek Biochemicals, Brea, CA) and ddTTP were tested by using the template-primer poly(rA)·oligo(dT). Inhibition by (−)-β-L-2′,3′-dideoxy-3′-thiacytidine triphosphate (3TCTP) (Moravek Biochemicals) was assayed by using single-strand M13mp18 DNA (New England Biolabs) hybridized to the sequencing primer −47 (New England Biolabs).

Processivity and Strand-Displacement Assays. The processivity assay was done as described (27, 28). For the strand displacement assay, an end-labeled PBS DNA oligonucleotide was hybridized to single-strand PPT-PBS sense DNA generated from PPT-PBS L-28 along with a 10-fold excess of four unlabeled DNA oligonucleotides. There was a 10-nt gap between the 3′ end of the labeled PBS oligonucleotide and the 5′ end of the first unlabeled DNA oligonucleotide; the four unlabeled DNA oligonucleotides hybridize to regions of the HIV-1 long terminal repeat upstream of the PBS region separated by 3-nt gaps.

The RT reaction conditions were similar to those described above, except that the buffer conditions included either 75 mM or 35 mM KCl and that mixture was incubated at 37°C for 30 min. The reactions were halted by extraction with phenol/chloroform and the nucleic acid was precipitated with isopropanol, the sample was resuspended in 10 μl of loading dye, and 4 μl was loaded on a 6% sequencing gel. After electrophoresis, the gel was exposed to x-ray film.

UTP Misincorporation. The UTP misincorporation assay was similar to the DNA-dependent DNA polymerase (DDDP) assays described above, except that the reaction conditions included 10 μM each of dATP, dGTP, and dCTP, 1 μM dTTP, and 2 μl of 800 Ci/mmol [α^{22}P]UTP (1 μM final concentration). After a 30-min incubation at 37°C, the reaction mixture was passed through Sephacryl S-200-h, extracted with phenol/chloroform, and precipitated with isopropanol. The sample was resuspended in 10 μl of loading dye and 4 μl was loaded on a 6% sequencing gel. After electrophoresis, the gel was exposed to x-ray film.

dGTP and GTP Kinetics. The dGTP kinetic assay was done as described (26, 28) except that 20 μl of 1.0 mM dGTP solution was added to 177 μl of ddH_2O and 3 μl of 800 Ci/mmol [α^{32}P]dGTP to make a final concentration of 100 μM. The GTP kinetic assay was done in a similar manner except that radioactive 100 μM GTP mixture was made by mixing 20 μl of 1 mM GTP with 170 μl of ddH_2O and 10 μl of 3,000 Ci/mmol [α^{32}P]GTP. The reaction time also was increased from 10 min to 30 min because GTP is incorporated less efficiently than dGTP.

Results

Polymerase Assays. In the RDDP assay, RT variants Tyr-115–Phe and Phe-116–Tyr had RDDP activity slightly higher than that of wild-type RT whereas the RDDP activity of Tyr-115–Val was approximately 75% the level of wild-type RT (Fig. 1*A*). These results are in agreement with the results we obtained with p66 homodimers (7, 8). RDDP activity also was measured with a PPT DNA oligonucleotide hybridized to antisense RNA and with a PBS DNA oligonucleotide hybridized to sense RNA. As shown in Fig. 1*A*, the RT variants Tyr-115–Phe and Phe-116–Tyr had activity levels with these substrates, relative to wild-type HIV-1 RT, that were similar to their relative activity with poly(rC)·oligo(dG). However, Tyr-115–Val had a relative RDDP activity with the HIV-1 RNAs (both sense and antisense) that was significantly lower than the relative RDDP activity obtained with poly(rC)·oligo(dG). One possible explanation for this discrepancy is that Tyr-115–Val is able to use homopolymeric templates more efficiently than heteropolymeric templates. We tested this possibility by using another homopolymeric template: poly(rA)·oligo(dT). With poly(rA)·oligo(dT), the RT variants Tyr-115–Phe and Phe-116–Tyr had relative RDDP activities similar to the RDDP activity obtained with other substrates (Fig. 1*A*). Although Tyr-115–Val had a higher relative RDDP activity with poly(rA)·oligo(dT) than with either the sense or the antisense HIV-1 RNA templates, the relative RDDP activity was still significantly lower than that measured with poly(rC)·oligo(dG) (Fig. 1*A*).

Another possible explanation is that Tyr-115–Val might have a lower fidelity than wild-type RT and occasionally might incorporate an incorrect dNTP into the growing primer strand while using the heteropolymeric template. Such a misincorporation might cause the enzyme to stall at the misincorporated base. Only one nucleotide (dGTP) is present when poly(rC)·oligo(dG) is the template primer, so there would be neither misincorporation nor stalling. To test this possibility, we repeated the assay with poly(rC)·oligo(dG) except with equal amounts of all four dNTPs in the reaction mixture. The level of radioactive dGTP incorporated into poly(rC)·oligo(dG) remained the same whether or not the other dNTPs were present (data not shown), suggesting that misincorporation is not responsible for the differences in the relative RDDP activity obtained with poly(rC)·oligo(dG) and the other substrates.

The DDDP activity of the RT variant Tyr-115–Phe was similar to that of wild-type RT for all three substrates (Fig. 1*B*), whereas Phe-116–Tyr had relative DDDP activity levels that were higher than wild-type RT with two of the substrates (Fig. 1*B*). Tyr-115–Val was significantly impaired with all of the DDDP substrates when compared with wild-type RT (Fig. 1*B*).

Processivity. The results of the processivity assays with the PBS oligonucleotide and a sense HIV-1 DNA substrate are shown in Fig. 2*A*. The RT variant Tyr-115–Phe had the same relative processivity as wild-type RT whereas Phe-116–Tyr was more processive and Tyr-115–Val was significantly less processive relative to wild-type RT. In the RNA genome of HIV-1, there are two well-characterized secondary structures located in the R-U5 region: the Tar element (1) and the poly(A) hairpin (34). These secondary structure elements, which also should be present in single-strand sense DNA, might cause HIV-1 RT to pause and increase the

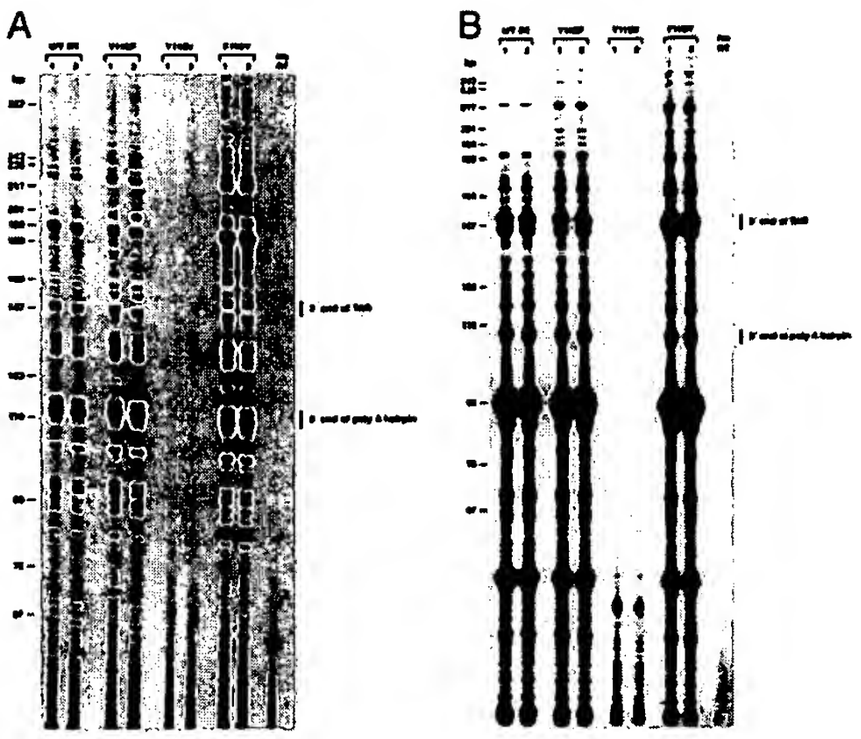


Fig. 2. Processivity assays with DNA and RNA templates. End-labeled primers were annealed to template DNA and RNA and extended in the presence of a cold trap that prevents multiple rounds of polymerization. (*A*) Processivity with the PBS DNA oligonucleotide annealed to sense PPT-PBS DNA. (*B*) Processivity with the PBS DNA oligonucleotide annealed to sense PPT-PBS RNA. The locations of the 3′ ends of two known secondary-structural elements, TAR (for a review, see ref. 1) and the poly(A) hairpin (34), are shown at the right. Duplicate assays, labeled 1 and 2, were done for each sample. WT, wild type.

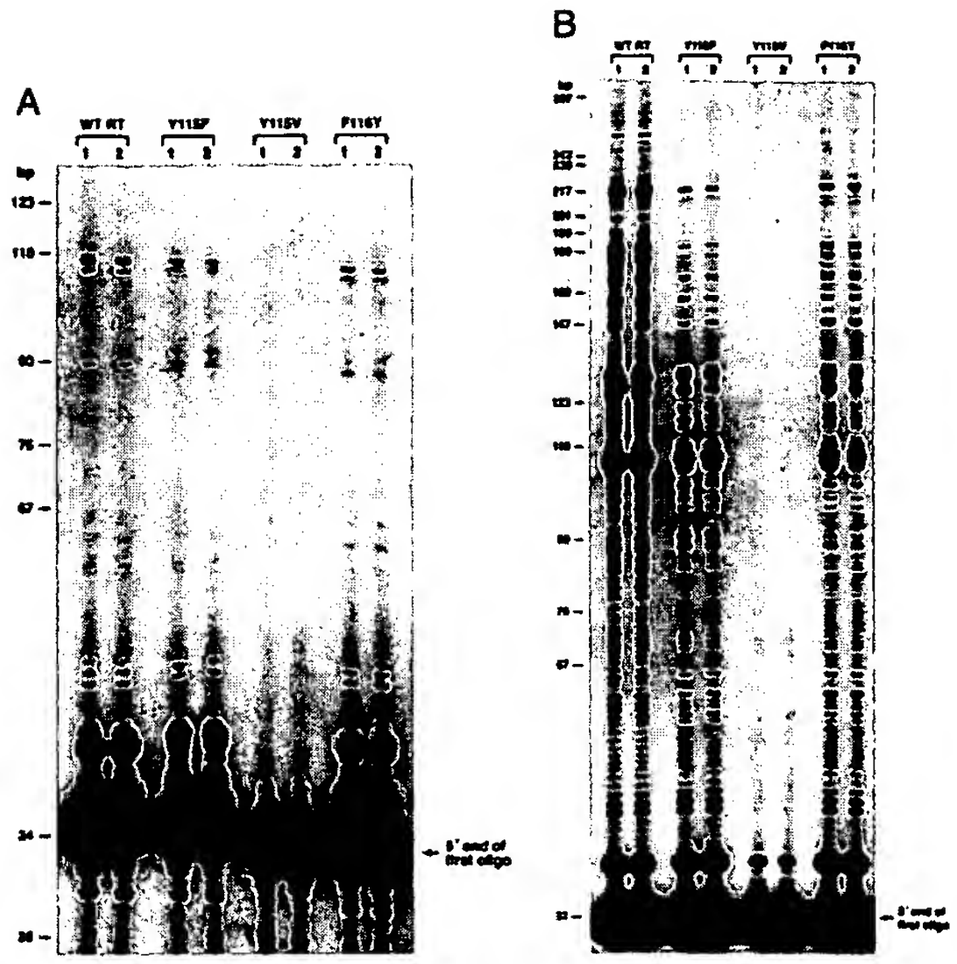


Fig. 3. Strand-displacement activity of wild-type (WT) and mutant HIV-1 RTs. The PBS DNA oligonucleotide was end-labeled with [γ^{32}P]ATP and annealed to sense single-strand DNA. A 10-fold excess of unlabeled DNA oligonucleotides also was annealed to the sense-strand DNA to provide the targets to be displaced. The RT polymerization reaction was done in the presence of two different concentrations of KCl: 75 mM KCl (*A*) and 35 mM KCl (*B*). The 5′ end of the first unlabeled target DNA oligonucleotide (marked at the right) indicates the starting point for strand displacement. Duplicate assays, labeled 1 and 2, were done for each sample.

probability of disassociation from the template primer. As shown in Fig. 2*A*, there are a series of extension products whose length corresponds to the 3′ end of the poly(A) hairpin, which is the element that the extending polymerase encounters first. There are some extension products that appear to terminate near the 3′ end of the TAR element; however, there are fewer of these products than those that correspond to the 3′ end of the poly(A) hairpin (Fig. 2*A*). Similar results were obtained when the processivity substrate was a DNA oligonucleotide annealed to single-strand M13mp18 DNA (data not shown). With a PPT oligonucleotide and an antisense HIV-1 DNA substrate, both Tyr-115–Phe and Phe-116–Tyr appeared to be slightly more processive and Tyr-115–Val was significantly less processive (data not shown).

The reported K_m for Tyr-115–Val for dTTP (62.7 μM) is significantly higher than wild-type RT (6.7 M), whereas Tyr-115–Phe has a reported K_m for dTTP of 3.0 μM (12). Although a dTTP concentration of 10 μM, which normally is used in the processivity assay, should be sufficient for wild-type RT, this concentration might be limiting for Tyr-115–Val because dTTP is present at a concentration approximately one-sixth of the reported K_m value. Therefore, we repeated the processivity assay by using a DNA oligonucleotide annealed to single-strand M13mp18 DNA but with the concentration of each nucleotide at 50 μM. The processivity for wild-type RT, Tyr-115–Phe, and Tyr-115–Val was not significantly different when the dNTP concentrations were 50 μM compared with the results with 10 μM concentrations (data not shown).

We also analyzed the processivity of the variant RTs on an RNA template. A DNA PBS oligonucleotide was hybridized to sense HIV-1 RNA (26). Again, Tyr-115–Phe and Phe-116–Tyr appear to have slightly higher processivity than the wild-type RT (Fig. 2*B*). Tyr-115–Val was significantly impaired compared with wild-type RT and the other RT variants. As described above, the sense RNA template would contain the TAR (1) and poly(A) hairpin (34) secondary structures. Interestingly, the RT extension products were somewhat different with RNA and DNA templates. As shown in Fig. 2*B*, the extension products corresponding to the 3′ end of the poly(A) hairpin, the region the RT encounters first, were not as intense with an RNA template as with a DNA template (Fig. 2*A*). The region of strongest pausing actually occurred before the poly(A) hairpin was reached (Fig. 2*A*) and there appeared to be more pausing at the 3′ end of the TAR element than with a single-strand sense DNA template.

Strand Displacement. We analyzed the strand-displacement activity of RT in the presence of both 75 mM KCl and 35 mM KCl, because Fuentes *et al.* (29) showed that 75 mM KCl (the concentration typically used in various RT assays) can inhibit strand displacement, presumably by stabilizing the nucleic acid duplex. They also showed that decreasing the concentration of KCl in the reaction increases the amount of strand displacement (14). As shown in Fig. 3*A*, 75 mM KCl inhibited strand-displacement activity by RT. Most of the PBS primers were extended only to the 5′ end of the first unlabeled oligonucleotide. The small amount of strand displacement by wild-type RT, Tyr-115–Phe, and Phe-116–Tyr extended no farther than 100 nt from the 5′ end of the first unlabeled DNA oligonucleotide. Tyr-115–Val did not appear to displace the DNA oligonucleotides to any measurable extent (Fig. 3*A*). In contrast, strand displacement was significantly increased in the presence of 35 mM KCl (Fig. 3*B*); wild-type RT extended the labeled PBS DNA oligonucleotide at least 200 nt into the double-strand substrate. Tyr-115–Phe and Phe-116–Tyr had significant levels of strand-displacement activity but did not appear to be as efficient as wild-type RT (Fig. 3*B*). Tyr-115–Val showed only a very low level of strand displacement activity even with the lower concentration of KCl.

Ribonucleotide Misincorporation. Gao *et al.* (21, 22) showed that the amino acid substitution Phe-155–Val in the MMLV RT allows the misincorporation of ribonucleotides into the growing primer strand. We tested whether mutants of HIV-1 RT with substitutions at Tyr-115, which is the amino acid equivalent to Phe-155 in the MMLV RT (30–32), would cause a similar misincorporation of ribonucleotides with a DNA:DNA template primer. Because

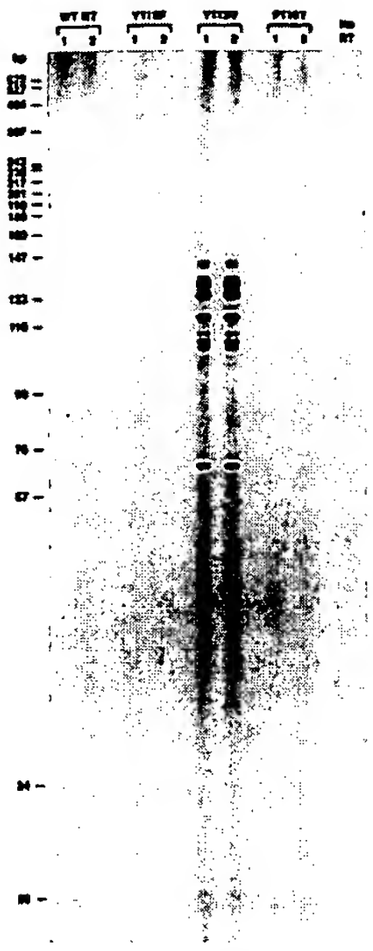


Fig. 4. UTP misincorporation by wild-type (WT) and mutant HIV-1 RTs. An unlabeled sequencing primer was annealed to single-strand M13mp18 DNA. RT polymerization was done with radioactive [α^{32}P]UTP in the reaction mixture. Extension products with UTP incorporation were radioactively labeled and visible on x-ray film. Duplicate assays, labeled 1 and 2, were done for each sample.

[α^{32}P]UTP was the only radioactive nucleotide in this assay, any UTP that was misincorporated would label the primer strand. Note that the concentrations of dTTP and UTP (1.0 μM) were equivalent in this reaction. As shown in Fig. 4, wild-type RT and Phe-116–Tyr did not produce labeled extension products. The level of UTP misincorporation, if any, was below the limit of detection in this assay. Tyr-115–Val had a high level of UTP misincorporation, producing labeled extension products ranging between 67 and 147 nt in length (Fig. 4). This result suggests that the Tyr-115–Val RT variant was able to continue extending the primer after misincorporation of the UTP. Surprisingly, Tyr-115–Phe also appeared to misincorporate UTP into the growing primer strand, albeit at a much lower level than Tyr-115–Val (Fig. 4). We repeated these experiments using [α^{32}P]ATP as the labeled ribonucleotide. Again, Tyr-115–Val incorporated the labeled ribonucleotide to a significant extent (data not shown). However, Tyr-115–Phe did not incorporate detectable amounts of ATP.

Kinetics of Ribonucleotide Incorporation. Using kinetic assays, we analyzed the ability of wild-type RT and Tyr-115–Val to use deoxynucleotides and ribonucleotides. Wild-type RT showed a V_{max} with dGTP of 38.8 ± 0.9 pmol incorporated/min and a K_m of 4.50 ± 0.2 μM, which is similar to our previous results (26, 28). In contrast to wild-type RT, Tyr-115–Val had a slightly lower V_{max} (28.1 ± 0.4 pmol incorporated/min) and a higher K_m (8.61 ± 0.3 μM) with dGTP (data not shown). The combination of a lower V_{max} and a higher K_m might explain why Tyr-115–Val is less processive than wild-type RT (Fig. 2*A*).

There were distinct differences between wild-type RT and Tyr-115–Val when GTP was the substrate. Wild-type RT had a somewhat higher V_{max} than did Tyr-115–Val (0.17 ± 0.01 pmol GTP incorporated/min versus 0.10 ± 0.03 pmol GTP incorporated/min); the V_{max} for either enzyme was much lower for GTP than dGTP. There was a larger difference in the K_m. Wild-type RT had a K_m of 23.1 ± 0.10 μM GTP; Tyr-115–Val had a K_m of 1.47 ± 0.12 μM GTP (data not shown).

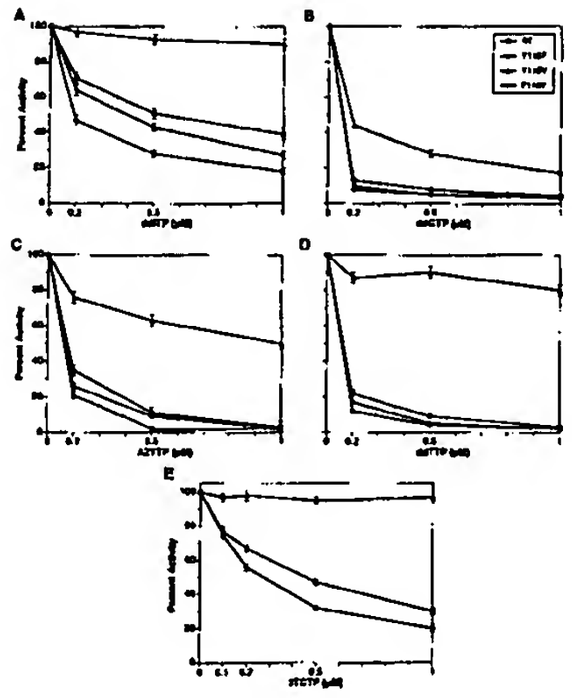


Fig. 5. Inhibition curves for various nucleoside analogs. The amount of radioactivity incorporated into the template-primer by wild-type and mutant HIV-1 RTs in the absence of inhibitor represents 100% enzyme activity. The amount of radioactivity incorporated at the various concentration of inhibitor is normalized to this value. The template primer for testing ddITP and ddGTP inhibition was poly(rC)·oligo(dG), whereas that for ddTTP and AZTTP was poly(rA)·oligo(dT). The template-primer for testing (−)-β-L-2′,3′-dideoxy-3′-thiacytidine triphosphate (3TCTP) inhibition was a sequencing primer annealed to single-strand M13mp18 DNA. When single-strand M13mp18 was used as a template with the −47 primer, the Tyr-115–Val mutant produced 36% as much DNA product as wild-type HIV-1 RT; with poly(rA)·oligo(dT) and poly(rC)·oligo(dT), the Tyr-115–Val mutant produced 50–75% as much DNA product.

Nucleoside Analog Resistance. Because Tyr-115–Val showed a decreased ability to discriminate between deoxyribose and ribose, we wanted to determine whether Tyr-115–Val could discriminate between normal dNTPs and various nucleoside analogs. We tested three dideoxyribonucleotide analogs as well as AZTTP and (−)-β-L-2′,3′-dideoxy-3′-thiacytidine triphosphate (3TLTP). As shown in Fig. 5, Tyr-115–Val was less sensitive to all of the nucleoside analogs than was wild-type RT. Tyr-115–Phe and Phe-116–Tyr appeared to be slightly more sensitive to the compounds than wild-type RT (Fig. 6). Tisdale *et al.* (16) showed that a virus containing the Tyr-115–Phe mutation is as susceptible as wild-type RT to several different nucleoside analogs. The exception is the carbocyclic nucleoside analog 1592U89; the resistance of the Tyr-115–Phe variant was about twice that of wild-type RT for this analog (16).

Discussion

Although the fidelity of HIV-1 RT is lower than that of some DNA polymerases, RT does select the proper dNTP to be added to the primer strand with considerable accuracy (approximately 1 error for every 10^4 bases polymerized) (see ref. 1 for a review). A dTTP residue has been modeled into the structure of HIV-1 RT in complex with double-strand DNA (5). A model for the structure of binding of a nucleoside triphosphate to MMLV RT also has been developed (31). Both models propose that the equivalent amino acids in the two structures (Tyr-115 in HIV-1 RT, Phe-155 in MMLV RT) interact with the ribose of the incoming triphosphate. The structure of a ternary complex of HIV-1 RT, double-strand DNA, and a bound dTTP (6) confirms these predictions and provides a more precise picture of the actual interactions of the protein and the triphosphate at the active site (Fig. 6). To explore the consequences of altering amino acids in the nucleotide binding pocket, we have analyzed the effects of the amino acid substitutions Tyr-115–Phe, Tyr-115–Val, and Phe-116–Tyr.

The effects of an amino acid substitution on HIV-1 RT can be subtle and may not be easily detected by using a simple polymerase assay with a homopolymeric template primer such as

nucleotides between the acceptor DNA and the newly synthesized plus strand is sufficient for either the WT RT or E478Q mutant enzyme to perform the plus-strand transfer. This region of overlap includes the 3′OH of the primer. Deletions of acceptor molecules which destroy the base pairing with the primer terminus block plus-strand transfer in vitro and in vivo (4, 54). Our results are consistent with in vivo results in which maintenance of 5 nucleotides of the PBS adjacent to U5 plus complementarity of only 3 bp at the site of polymerization was sufficient for extension of plus-strand DNA during the second template transfer (54). This result is also in agreement with the occurrence of retroviral recombination events. In vivo, low levels of strand transfer events that yield deletions have been characterized with minimal overlap (41). Studies have shown that there is an average of 1 aberrant strand transfer event per replication cycle (29).

Interestingly, WT HIV-1 RT was capable of performing specific RNase H cleavages on all of the substrates utilized. Truncation of the DNA or RNA portion did not affect the RNase H cleavage site. Rather, all of the substrates produced the identical −1 cleavage product. This may be due to the strong structural recognition of the tRNA mimic or to an alteration of the binding conformation which allows specific RNase H cleavages to occur. Previously, we have demonstrated that sequences within the first nine positions of the tRNA were important for cleavage and recognition by an isolated RNase H domain (49). These studies support the concept that the structure defined by the tRNA sequence possesses strong intrinsic signals that lead to its precise cleavage between the terminal ribo-A and ribo-C. This cleavage occurs even when a large portion of the $tRNA^{Lys,3}$ sequence has been deleted.

We have shown that many factors play a role in influencing RNase H activity. The in vitro system developed allows the biochemical analysis of a minimal system. In vivo, replication occurs within reverse transcription complexes (19) in the presence of additional viral proteins, including the nucleocapsid, which can influence these reactions. The use of an enzyme with defects in the RNase H domain provides insight into the requirements for polymerase-independent RNase H cleavages.

ACKNOWLEDGMENTS

This work was supported by NIH (grants RO1GM51151 and 1RO1 CA90174).

We thank Jeff Culp and Christine DeBouch for the generous gift of HIV-1 RT and Stuart J. F. Le Grice for the gift of E478Q RT and HIV-1 RT. We thank Millie Georgiadis for assistance in molecular modeling.

REFERENCES

1. **Arnold, E., A. Jacobo-Molina, R. G. Nanni, R. L. Williams, X. Lu, J. Ding, J. Arthur, D. Clark, A. Zhang, A. L. Ferris, P. Clark, A. Hizi, and S. H. Hughes.** 1992. Structure of HIV-1 reverse transcriptase/DNA complex a 7A resolution showing active site locations. Nature 357:85–89.
2. **Arts, E. J., and S. F. L. Grice.** 1998. Interaction of retroviral reverse transcriptase with template-primer duplex during replication. Prog. Nucleic Acid Res. Mol. Biol. 58:339–393.
3. **Arts, E. J., S. R. Stetor, S. Li, J. W. Rausch, K. J. Howard, B. Ehresmann, T. W. North, B. M. Wohrl, R. S. Goody, M. A. Wainberg, and S. F. J. LeGrice.** 1996. Initiation of (−) strand DNA synthesis from $tRNA^{Lys,3}$ on lentiviral RNAs: implications of specific HIV-1 RNA-$tRNA^{Lys,3}$ interactions inhibiting primer utilization by retroviral reverse transcriptases. Proc. Natl. Acad. Sci. USA 93:10063–10068.
4. **Auxilien, S., G. Keith, S. F. J. Le Grice, and J.-L. Darlix.** 1999. Role of post-transcriptional modifications of primer $tRNA^{Lys,3}$ in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. J. Biol. Chem. 274:4412–4420.
5. **Beard, W. A., K. Bebenek, T. A. Darden, L. Li, R. Prasad, T. A. Kunkel, and S. H. Wilson.** 1998. Vertical-scanning mutagenesis of a critical tryptophan in the minor groove binding track of HIV-1 reverse transcriptase. J. Biol. Chem. 273:30435–30442.
6. **Bebenek, K., W. A. Beard, J. R. Casas-Finet, H. R. Kim, T. A. Darden, S. H. Wilson, and T. A. Kunkel.** 1995. Reduced frameshift fidelity and processivity of HIV-1 reverse transcriptase mutants containing alanine substitutions in helix H of the thumb subdomain. J. Biol. Chem. 270:19516–19523.
7. **Ben-Artzi, H., J. Shemesh, E. Zeelon, B. Amit, L. Kleiman, M. Gorecki, and A. Panet.** 1996. Molecular analysis of the second template switch during reverse transcription of the HIV RNA template. Biochemistry 35:10549–10557.
8. **Blain, S. W., and S. P. Goff.** 1995. Effects on DNA synthesis and translocation caused by mutations in the RNase H domain of Moloney murine leukemia virus reverse transcriptase. J. Virol. 69:4440–4452.
9. **Cameron, C. E.** 1997. Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand transfer efficiency are suppressed by HIV nucleocapsid protein. Biochemistry 94:6700–6705.
10. **Cirino, N. M., C. E. Cameron, J. S. Smith, J. W. Rausch, M. J. Roth, S. J. Benkovic, and S. F. J. Le Grice.** 1995. Divalent cation modulation of ribonuclease functions of human immunodeficiency virus reverse transcriptase. Biochemistry 34:9936–9943.
11. **Das, K., J. Ding, Y. Hsiou, A. D. J. Clark, H. Moereels, L. Koymans, K. Andries, R. Pauwel, P. A. J. Janssen, P. L. Boyer, P. Clark, R. H. J. Smith, M. B. K. Smith, C. J. Michejda, S. H. Hughes, and E. Arnold.** 1996. Crystal structure of 8-Cl and 9-Cl TIBO complexed with wild type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. J. Mol. Biol. 264:1085–1100.
12. **DeStefano, J.** 1995. The orientation of human immunodeficiency virus reverse transcriptase on nucleic acid hybrids. Nucleic Acids Res. 23:3901–3908.
13. **DeStefano, J. J., W. Wu, J. Seehra, J. M. Coy, D. Laston, E. Albone, P. J. Fay, and R. A. Bambara.** 1994. Characterization of an RNase H defective mutant of human immunodeficiency virus type 1 reverse transcriptase having an aspartate to asparagine change at position 498. Biochim. Biophys. Acta 1219:380–388.
14. **DiMarzoVeronese, F., T. D. Copeland, A. L. D. Vico, R. Rahman, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1986. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-II/LAV. Science 231:1289–1291.
15. **Ding, J., K. Das, H. Moerells, L. Koymans, K. Andries, P. A. Janseen, S. H. Hughes, and E. Arnold.** 1995. Structure of HIV-1 RT/TIBO R86183 reveals similarity in the binding of diverse nonnucleoside inhibitors. Nat. Struct. Biol. 2:407–415.
16. **Ding, J., K. Das, C. Tantillo, W. Zhang, A. D. Clark, S. Jessen Jr., X. Lu, Y. Hsiou, A. Jacobo-Molina, K. Andries, R. Pawels, H. Moereels, L. Koymans, P. A. J. Janssen, and R. H. Smith, Jr.** 1995. Structure of HIV-1 reverse transcriptase in a complex with the nonnucleoside inhibitor α-APA R95845 at 2.8A resolution. Structure 3:365–379.
17. **Dudding, L. R., N. C. Nkabinde, and V. Mizrahi.** 1991. Analysis of the RNA- and DNA-dependent DNA polymerase activities of point mutants of HIV-1 reverse transcriptase lacking ribonuclease H activity. Biochemistry 30:10498–10506.
18. **Fan, N., K. B. Rank, D. E. Slade, S. M. Poppe, D. B. Evans, L. A. Kopta, R. A. Olmsted, R. C. Thomas, W. G. Tarpley, and S. K. Sharma.** 1996. A drug resistance mutation in the inhibitor binding pocket of human immunodeficiency virus type 1 reverse transcriptase impairs DNA synthesis and RNA degradation. Biochemistry 35:9737–9745.
19. **Fassati, A., and S. P. Goff.** 1999. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. J. Virol. 73:8919–8925.
20. **Furfine, E. S., and J. E. Reardon.** 1991. Reverse transcriptase-RNase H from the human immunodeficiency virus. J. Biol. Chem. 266:406–412.
21. **Gao, H.-Q., P. L. Boyer, E. Arnold, and S. H. Hughes.** 1998. Effects of mutations in the polymerase domain on the polymerase, RNase H and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase. J. Mol. Biol. 277:559–572.
22. **Gao, H.-Q., S. G. Sarfianos, E. Arnold, and S. H. Hughes.** 1999. Similarities and differences in the RNase H activities of human immunodeficiency virus type 1 reverse transcriptase and Moloney murine leukemia virus reverse transcriptase. J. Mol. Biol. 294:1097–1113.
23. **Ghosh, M., K. J. Howard, C. E. Cameron, S. J. Benkovic, S. H. Hughes, and S. F. J. LeGrice.** 1995. Truncating alpha-helix E′ of p66 human immunodeficiency virus reverse transcriptase modulates RNase H function and impairs DNA strand transfer. J. Biol. Chem. 270:7058–7976.
24. **Huang, H., R. Chopra, G. L. Verdine, and S. C. Harrison.** 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282:1669–1675.
25. **Isel, C., J. M. Lanchy, S. F. L. Grice, C. Ehresmann, B. Ehresmann, and R. Marquet.** 1996. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer $tRNA^{3Lys}$. EMBO J. 1996. 15:917–924.
26. **Isel, C., R. Marquet, G. Keith, C. Ehresmann, and B. Ehresmann.** 1993. Modified nucleotides of $tRNA^{(3Lys)}$ modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. J. Biol. Chem. 268:25269–25272.
27. **Isel, C., E. Westhof, C. Massire, S. F. Le Grice, B. Ehresmann, C. Ehres-**

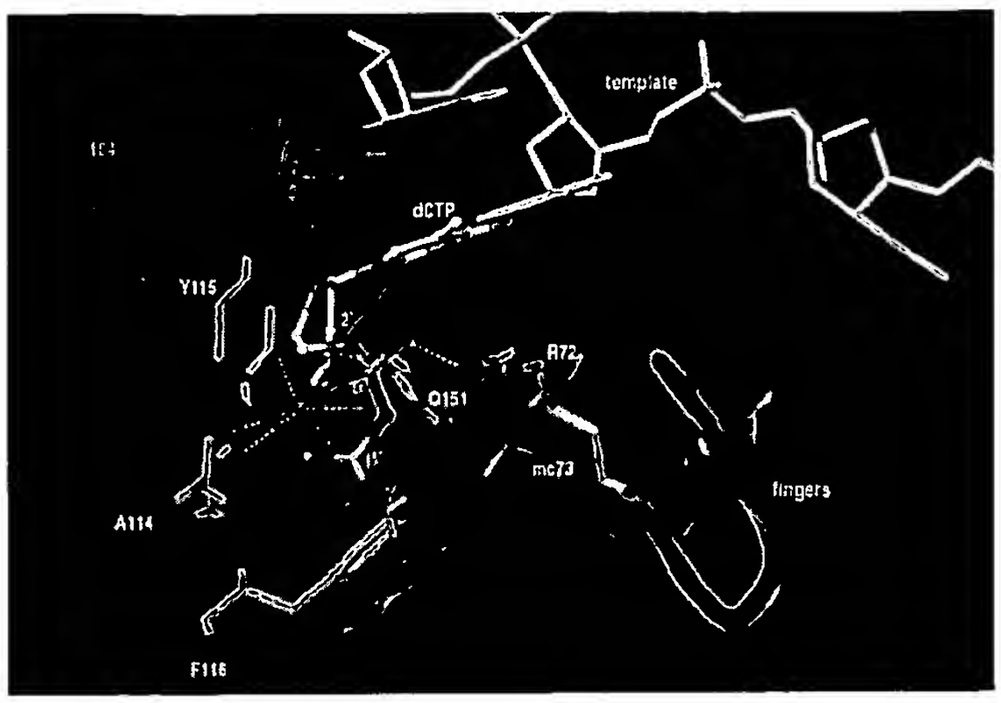


Fig. 6. View of the polymerase active site of HIV-1 RT. The sugar ring of the incoming dCTP is positioned between the hydrophobic side chains of Tyr-115 and Ala-114 at the back (Van der Waals spheres in purple), and Glu-151 and Arg-72 in front (Van der Waals spheres in cyan). Phe-116, although too far away to interact directly with the sugar of the dCTP, forms the floor of the 3′ OH binding pocket (Van der Waals spheres in purple). Yellow lines indicate interactions of the 3′ OH of dCTP with the ring of Tyr-115, the main-chain NH of Tyr-115, the main-chain NH of Ala-114, β-phosphate of dCTP, and CO of Glu-151. Bridging interactions of the NH group of Glu-151 with the guanidinium of Arg-72 and the main-chain CO of Lys-73 are shown as green dotted lines. The interactions of the Tyr-115 ring and the Glu-151 side chain with the 2′ carbon of dCTP are shown as red dotted lines.

poly(rC)·oligo(dG). Using HIV-1 genome-based templates, Tyr-115–Phe is similar to wild-type RT in the RDDP, DDDP, and processivity assays (Figs. 1 *A* and *B* and 2 *A* and *B*). However, it is clear in the strand displacement assay that the RT variant Tyr-115–Phe is impaired relative to the wild-type enzyme (Fig. 3). As will be discussed below, Tyr-115–Phe also might have a decreased ability to select dNTPs over NTPs. In contrast, the RT variant Phe-116–Tyr appears to be better than wild-type RT in most of the RDDP and DDDP assays (Fig. 1 *A* and *B*) and is more processive than wild-type RT (Fig. 2*A* and *B*). However, this variant also is impaired in the strand displacement assay relative to wild-type RT (Fig. 3*A*).

The effects of Tyr-115–Val amino acid substitution are not subtle, however. With the exception of the template primer poly(rC)·oligo(dG), this enzyme is significantly impaired in both the RDDP and DDDP assays compared with wild-type RT (Fig. 1 *A* and *B*). The Tyr-115–Val variant is also significantly less processive than wild-type RT with a variety of template-primers (Fig. 2 *A* and *B*). Tyr-115–Val has been reported to have a much higher K_m for dTTP (62.7 μM) than does wild-type RT (6.7 μM) (12). We found that Tyr-115–Val has a higher K_m for dGTP than does wild-type RT (8.61 μM dGTP versus 4.50 μM) and a lower V_{max} (28.1 pmol dGTP incorporated/min versus 38.8 pmol incorporated/min for wild-type RT). However, the K_m we measured for dGTP is significantly lower than the reported value for dTTP. Taken together, the available data suggest not only that the Tyr-115–Val RT variant binds the incoming dNTP more weakly than wild-type RT but that it binds the incoming dNTP improperly: the difference in the V_{max} shows that increasing the concentration of dNTPs did not completely compensate for the improper binding of the dNTP.

Mispositioning of the incoming nucleotide also might allow the RT variant Tyr-115–Val to incorporate NTPs (Fig. 4). The equivalent Phe-155–Val mutation in MMLV RT causes a loss of discrimination between dNTPs and NTPs rather than a complete switch of substrate specificity (21, 33). Wild-type MMLV RT has a K_m for UTP that is much higher than the K_m for dTTP (443 μM versus 9.2 μM). There is a similar effect on V_{max}. Both wild-type MMLV RT and the Phe-155–Val mutant have similar V_{max} values for dTTP (0.38 and 0.41 μmol/min per mg, respectively). The rate for dTTP incorporation is much higher than for UTP (1.17 versus 5.45 μmol/min per mg). For the Phe-155–Val variant, the K_m values are similar (4.34 μM versus 13.7 μM for the wild-type RT) (21). Our data suggest that the Tyr-115–Val variant of HIV-1 RT also has lost substrate specificity. However, there are important differences between the results obtained with the Phe-155–Val MMLV RT variant and those obtained with the Tyr-115–Val HIV-1 RT variant. As described above, Tyr-115–Val has been reported to have a significantly higher K_m for dTTP (62.7 μM) than wild-type HIV-1 RT (6.7 μM) (12). The difference in the dTTP K_m values between wild-type MMLV RT and the MMLV RT variant is much smaller (21).

We measured the K_m and V_{max} values for wild-type HIV-1 RT and the Tyr-115–Val variant with both dGTP and GTP. In contrast to wild-type MMLV RT, wild-type HIV-1 RT did not have a large difference in the K_m values between dGTP and GTP. The K_m for dGTP was 4.50 μM whereas the K_m for GTP was 23.1 μM. The most significant difference was in the V_{max} for wild-type RT, which was 38.8 pmol incorporated/min for dGTP versus only 0.17 pmol incorporated/min for GTP (data not shown). This result suggests that wild-type HIV-1 RT can bind GTP, but does so in a fashion that is unfavorable for polymerization. Interestingly, Tyr-115–Val had a somewhat lower V_{max} with GTP than did wild-type RT (data not shown), yet in the UTP misincorporation assay, Tyr-115–Val clearly incorporated UTP into the extended primer whereas no incorporation was detected for wild-type HIV-1 RT (Fig. 4).

The answer to this apparent paradox might be found in the K_m values. The Tyr-115–Val mutant had a low K_m value (1.47 μM), whereas the wild-type RT had a significantly higher K_m (23.1 μM). The level of [α^{32}P]UTP in the UTP misincorporation assay was low (1 μM) and was equal to the level of the preferred substrate (1 μM dTTP). If Tyr-115–Val also had a lower K_m for UTP relative to wild-type RT, the results of the UTP misincorporation might reflect an ability of Tyr-115–Val to bind NTPs at the polymerase active site better than wild-type RT, rather than an increase in the ability to (mis)incorporate NTPs.

It has been proposed that the presence of a phenyl ring at residue 155 of MMLV RT acts directly as a steric gate that distinguishes between a 2′ H group of a dNTP and a 2′ OH group of an NTP (22, 28, 31). In this model, the Phe-155–Val mutant replaces the bulky aromatic side chain with the smaller side group of valine and removes the steric gate. The structure of a ternary complex of HIV-1 RT, double-strand DNA, and a dTTP (Fig. 6), although different from the proposed model in structural detail, suggests that the corresponding amino acid Tyr-115 plays a role in forming a steric gate in concert with Q151 (5). As shown in Fig. 6, the 2′ and 3′ positions of the incoming dNTP are flanked on one side by the Tyr-115 aromatic side chain and on the other side by the Q151 side chain. The 2′ position of the dNTP is closer to the aliphatic side chain and the main-chain carbonyl oxygen of Q151 (3.4 and 3.0 Å) than to the side chain of Tyr-115 (3.5 Å). Nevertheless, the sugar puckering of the dNTP directs a 2′ OH to the ring of Tyr-115 rather than to the Q151 side chain. In the Tyr-115Val mutant, the aromatic side chain is replaced by a shorter, more rigid, branched residue. This side chain would be expected to move toward the dNTP binding site because of unfavorable interactions with the main-chain carbonyl oxygen of residue 149, a movement that should permit the introduction of an OH group at the 2′ sugar position.

Although the mutant Tyr-115–Val was able to incorporate UTP, ATP, and GTP more efficiently than wild-type RT, the other properties of Tyr-115–Val suggest that the amino acid at position 115 has additional roles beyond that of participating in a simple steric gate. Tyr-115–Val was clearly defective compared with wild-type HIV-1 RT in most of our assays. This mutant had a much lower processivity activity with both RNA and DNA templates than did wild-type RT, and it was unable to cause strand displacement under any of the conditions we tested. However, Tyr-115–Val RT showed

a high degree of resistance to all of the nucleoside analogs that we tested. In the absence of a three-dimensional structure of the Tyr-115–Val RT variant with a bound nucleotide, it is unclear how the variant is able to select the proper dNTP over a nucleoside analog more effectively than wild-type HIV-1 RT. However, if the Tyr-115–Val variant inappropriately positions the incoming dNTP, it also might bind the nucleoside analog incorrectly, suggesting that there are steric constraints that help the mutant enzyme discriminate against certain modified sugars. Although the degree of resistance to nucleoside analogs is significant, the Tyr-115–Val mutation has not been found in HIV-1 patients treated with nucleoside analogs. This is presumably a result of the fact that mutations like Tyr-115–Val cause sufficient distortion of the polymerase active site such that HIV-1 variants carrying these mutations cannot be selected. In support of this idea, an MMLV RT variant with an Phe-155–Val mutation does not replicate (22).

Tyr-115–Val is more efficient at extending a mismatched base pair than is wild-type HIV RT (12). It should be remembered that this assay does not directly measure the fidelity; rather, it measures how well the variants can incorporate dNTPs with a less-than-optimal template primer configuration at the active site. The incoming dNTP can be considered to have two components: a base and a ribose phosphate. The overall fidelity of the enzyme depends primarily on how well the base of the incoming dNTP is matched to the complementary base on the template strand. The mispair extension assay primarily involves the ribose phosphate and how well the enzyme can add an incoming dNTP to the mispaired base at the 3′ end of the primer strand. As discussed above, the Tyr-115–Val variant may improperly position the incoming dNTP. This mispositioning of the dNTP might allow the Tyr-115–Val variant to extend a mismatched pair more efficiently than wild-type HIV-1 RT, but this result does not necessarily mean the enzyme has a lower fidelity than the wild-type enzyme. Although the ability to extend a primer to which the wrong base already has been added is a component of fidelity, this event is secondary to the misincorporation of a dNTP.

It has been suggested that an MMLV containing the Phe-155–Val mutation in the RT coding region is defective in its replication because ribonucleotides, once incorporated into the growing DNA strand, act as inhibitors of DNA synthesis (22). However, the incorporation of ribonucleotides does not appear to block DNA synthesis by the HIV-1 RT variants Tyr-115–Phe and Tyr-115–Val (Fig. 4). It should be noted that the extension products for both Tyr-115–Val and Tyr-115–Phe in the UTP misincorporation assay are relatively large. The major extension products for Tyr-115–Val range in size between 67 and 147 nt, and the extension products for Tyr-115–Phe are more than 300 nt in size. This result suggests that both Tyr-115–Val and Tyr-115–Phe are able to extend the primer strand after the misincorporation of UTP.

In summary, the amino acid residue Tyr-115 of HIV-1 RT has an important role in the selection of the proper nucleotide during polymerization. Although the amino acid substitution Tyr-115–Phe has little effect on the enzyme, the mutation Tyr-115–Val has deleterious effects on polymerization. The equivalent positions in other polymerases also appear to function in the selection of the proper nucleotide and in the rejection of inappropriate substrates (for a review, see ref. 30). As more is learned about the nucleotide binding region(s) of polymerases, it may be possible to engineer polymerases with novel substrate specificities (33).

We are most grateful to Pat Clark and Peter Frank for enzyme purification, A. Arthur for expert editorial assistance, and H. Marusiodis for help in preparing this paper. The research in S.H.H.'s laboratory was sponsored by the National Cancer Institute, Department of Health and Human Services, under contract with Advanced BioScience Laboratories, and by the National Institute of General Medical Sciences. The research in E.A.'s laboratory was supported by National Institutes of Health Grants GM 56609 and AI 27690 (to E.A.) and National Institutes of Health National Research Service Award Postdoctoral Fellowship AI 09578 (to S.G.S.).

1. Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997) *Retroviruses* (Cold Spring Harbor Lab. Press, Plainview, NY).
2. De Clercq, E. (1994) *Biochem. Pharmacol.* **47,** 155–169.
3. Mellors, J. W., Larder, B. A. & Schinazi, R. G. (1995) *Int. Antiviral News* **3,** 8–13.
4. Richman, D. D. (1993) *Antimicrob. Agents Chemother.* **37,** 1207–1213.
5. Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S. H. & Arnold, E. (1998) *J. Mol. Biol.* **284,** 1095–1111.
6. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998) *Science* **282,** 1669–1675.
7. Boyer, P. L., Ferris, A. L. & Hughes, S. H. (1992) *J. Virol.* **66,** 1031–1039.
8. Boyer, P. L., Ferris, A. L., Clark, P., Whitmer, J., Frank, P., Tantillo, C., Arnold, E. & Hughes, S. H. (1994) *J. Mol. Biol.* **243,** 472–483.
9. Larder, B. A., Kemp, S. D. & Purifoy, D. J. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4803–4807.
10. Lowe, D. M., Parmar, V., Kemp. S. D. & Larder, B. A. (1991) *FEBS Lett.* **282,** 231–234.
11. Martín-Hernández, A., Domingo, E. & Menèndez-Arias, L. (1996) *EMBO J.* **15,** 4434–4442.
12. Martín-Hernández, A. M., Gutièrrez-Rivas, M., Domingo, E. & Menèndez-Arias, L. (1997) *Nucleic Acids Res.* **25,** 1383–1389.
13. Wrobel, J. A., Chao, S.-F., Conrad, M. J., Merker, J. D., Swanstrom, R., Pielak, G. J. & Hutchison, C. A., III (1998) *Proc. Natl. Acad. Sci. USA* **95,** 638–645.
14. Harris, D., Kaushik, N., Pandey, P. K., Yadav, P. N. S. & Pandey, V. N. (1998) *J. Biol. Chem.* **273,** 33624–33634.
15. Vandamme, A.-M., Van Vaerenbergh, K. & De Clercq, E. (1998) *Antiviral Chem. Chemother.* **9,** 187–203.
16. Tisdale, M., Alnadaf, T. & Cousens, D. (1997) *Antimicrob. Agents Chemother.* **41,** 1094–1098.
17. Rezende, L. F., Curr, K., Ueno, T., Mitsuya, H. & Prasad, V. R. (1998) *J. Virol.* **72,** 2890–2895.
18. Schafer, R. W., Kozal, M. J., Winters, M. A., Iverson, A. K. N., Katzenstein, D. A., Ragni, M. V., Meyer, W. A., III, Gupta, P., Rasheed, S., Coombs, R., *et al.* (1994) *J. Infect. Dis.* **169,** 722–729.
19. Shirasaka, T., Kavlick, M. F., Ueno, T., Gao, W.-Y., Kojima, E., Alcaide, M. L., Chokekijchai, S., Roy, B. M., Arnold, E., Yarchoan, R. & Mitsuya, H. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2398–2402.
20. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989) *EMBO J.* **8,** 3867–3874.
21. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A. & Goff, S. P. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 407–411.
22. Gao, G. & Goff, S. P. (1998) *J. Virol.* **72,** 5905–5911.
23. Boyer, P. L., Ferris, A. L. & Hughes, S. H. (1992) *J. Virol.* **66,** 7533–7537.
24. Boyer, P. L., Tantillo, C., Jacobo-Molina, A., Nanni, R. G., Ding, J., Arnold, E. & Hughes, S. H. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4882–4886.
25. Adachi, A., Gendelman, H. E., Koenig, S., Folks, T., Wiley, R. A., Rabson, A. & Martin, M. A. (1986) *J. Virol.* **59,** 284–291.
26. Boyer, P. L., Liszienicz, J., Lori, F. & Hughes, S. H. (1999) *J. Mol. Biol.* **286,** 995–1008.
27. Boyer, P. L. & Hughes, S. H. (1995) *Antimicrob. Agents Chemother.* **39,** 1624–1628.
28. Boyer, P. L., Gao, H.-Q. & Hughes, S. H. (1998) *Antimicrob. Agents Chemother.* **42,** 447–452.
29. Fuentes, G. M., Rodríguez, L., Palaniappan, C., Fay, P. J. & Bambara, R. A. (1996) *J. Biol. Chem.* **271,** 1966–1971.
30. Barber, A., Hizi, A., Maizel, J. V., Jr. & Hughes, S. H. (1990) *AIDS Res. Human Retrovir.* **6,** 1061–1072.
31. Georgiadis, M. M., Jessen, S. M., Ogata, C. M., Telesnitsky, A., Goff, S. P. & Hendrickson, W. A. (1995) *Structure (London)* **3,** 879–892.
32. Johnson, M. S., McClure, M. A., Feng, D. F., Gray, J. & Doolittle, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7648–7652.
33. Joyce, C. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1619–1622.
34. Das, A. T., Klaver, B., Klasens, B. I. F., van Wamel, J. L. B. & Berkhout, B. (1997) *J. Virol.* **71,** 2346–2356.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ❑ **BLACK BORDERS**
- ❑ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ❑ **FADED TEXT OR DRAWING**
- ❑ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ❑ **SKEWED/SLANTED IMAGES**
- ☑ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☑ **GRAY SCALE DOCUMENTS**
- ☑ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ❑ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ❑ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.